N-63-2-5

RADC-TDR-62-533

15 January 1963

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS

Report Prepared by: Hans J. Queisser

SHOCKLEY TRANSISTOR
Unit of Clevite Transistor
Division of Clevite Corporation
1801 Page Mill Road
Palo Alto, California

Final Report

Contract AF 30(602)-2556

Prepared for

Rome Air Development Center
Air Force Systems Command
United States Air Force
Griffiss Air Force Base
New York

RADC-TDR-62-533

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS

Report Prepared by: Hans J. Queisser

SHOCKLEY TRANSISTOR
Unit of Clevite Transistor
Division of Clevite Corporation
1801 Page Mill Road
Palo Alto, California

Final Report

Contract AF 30(602)-2556
Project 5519
Task 45155

Prepared for

Rome Air Development Center
Air Force Systems Command
United States Air Force
Griffiss Air Force Base
New York

PREFACE

The work reported here represents one of several investigations being sponsored by the Rome Air Development Center under a broad reliability physics program. The general objective of this broad program is to detect, identify, and measure basic failure mechanisms in several classes of materials and devices, and apply such information to the assessment, prediction, and improvement of the reliability of electron devices.

The specific objective of the work reported in this Final Report was to investigate failure mechanisms due to bulk defects in silicon semiconductor devices.

One bulk defect of particular interest is the dislocation. Dislocations were investigated with specific attention directed at ways in which they can degrade devices. Under this contract, several experimental techniques have been developed to investigate the diffusion enhancement properties of dislocations.

Strong evidence against the dangling bond model for dislocations in silicon is presented and discussed. The results explain why dislocations in silicon devices do not directly influence device properties. Electrical effects of dislocations in semiconductor junction devices are determined not only by impurity atmospheres and preferential SiO_2 precipitation, but perhaps also by metals or carbon and their oxygen compounds.

Effects of twins, a structural defect very common in silicon, are exhaustively investigated. Twins are classified according to their energies. High energy (incoherent) twins can induce softness in p-n junction by precipitation of impurities, but these effects are weaker than expected. Low energy (coherent) twin boundaries have no unusual diffusion behavior nor do they influence p-n junction characteristics.

Epitaxial silicon layers, now commonly used in high-reliability transistor structures, are shown to be especially susceptible to stacking fault defects. Stacking faults originate mainly from oxygen impurities on an imperfectly prepared substrate. It is shown that at the points of intersection of stacking faults, local regions of low breakdown strength occur, resulting in serious limitation of device capability and reliability.

The problem of secondary or thermal breakdown in high power transistors has been investigated both theoretically and experimentally. It is shown that thermal breakdown can be initiated by lateral instability which channels a disproportionate share of the current through a small area in the semiconducting material, continuously heating the material until, ultimately, the material melts and the device is destroyed. Theoretically, the mathematics of the instability is structured as a function of the thermal, electrical, and geometric properties of the material. It is shown that a 10 per cent change in local current may result from a 1°C local temperature change.

Experimental studies on a high power interdigitated silicon structure have confirmed the elements of the theory. Sensitive regions in the semiconductor material have been detected which are especially sensitive to thermal instability. The nature of these regions is not yet understood. Definitive experiments concerning such defects should be undertaken.

The studies of thermal breakdown and lateral instability have been applied by the contractor in the production of a higher power Shockley four-layer diode for RADC as a pulse modulator switch contract AF 30 (602)-2811. Thus the reliability physics program can be considered to have produced useful results in a relatively short time.

Altogether, this TDR represents a considerable synthesis of bulk problems in silicon semiconductor devices. Dislocations, grain boundaries, twin boundaries, and stacking faults have been exhaustively studied, and many of their effects pertinent to device reliability should now be considered to be known. The work related to secondary breakdown should be continued both theoretically and experimentally. Theoretical findings should be applied to improve existing structures, and new high power structures should be designed specifically to mitigate the failure mechanism caused by instabilities.

The completeness of the work on bulk failure mechanisms would seem to logically direct attention to increasing awareness of the importance of semiconductor surfaces and interfaces in device failure. With

recent advances in surface chemistry and surface physics, it is felt that significant contributions can now be made to the physics of failure program in this area.

The investigation conducted under this contract is considered to have successfully achieved the objectives initially specified and has clearly indicated areas for future emphasis. A considerable portion of the results has been published separately in the open scientific literature. The significance of the results to device reliability lies in the knowledge of basic defect-environment reactions which have been brought to light. This report represents a very worthwhile contribution to the physics of failure program.

S/ ALFRED L. TAMBURRINO
PROJECT ENGINEER
Applied Research Laboratory
Rome Air Development Center

Foreword

This Final Report covers the work done under Contract AF30 (602)-2556, Project number 5519, task 45155 during the period from July 1, 1961 to September 30, 1962. The contract is a part of the "Physics of Failure" Program carried out by Rome Air Development Center.

The report consists of two major parts. The first part covers work on lattice defects in silicon and their influence upon performance and reliability of silicon devices. The second part describes a lateral thermal instability in silicon power transistors.

The report describes the joint efforts of several investigators of this organization Participating in the studies on lattice defects were R. Finch, W. Hooper, P. van Loon, H. Queisser, and R. Regehr. The pedestal method was applied to silicon bicrystal growth by R. Gereth and K. Alvin; their work was sponsored in part by this contract. The research on stacking faults was done in cooperation with Profs. J. Washburn and G. Thomas of the University of California at Berkeley. (Lawrence Radiation Laboratory). The transmission electron microscopy was carried out by them under sponsorship by the Atomic Energy Commission.

The major portions of the second part of this report were contributed by W. Shockley, R. M. Scarlett, R. Haitz and W. Hooper. The initial work on this problem of lateral instability in power transistors was carried out without government sponsorship. The later phases of these investigations were supported under this contract. The entire work is described in this report for the sake of completeness.

H. J. Queisser is the editor of this report. He wishes to offer his thanks to all the colleagues mentioned above for their assistance, especially to Dr. W. Shockley for his aid and encouragement.

Abstract

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, twin boundaries, and stacking faults in epitaxial layers are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown.

The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown" failure mechanisms in transistors.

Publications Covering Work Performed under this Contract

1. "Stacking Faults in Epitaxial Silicon"
   H. J. Queisser, R. H. Finch and J. Washburn
   J. Appl. Phys. 33, 1536 (1962)

2. "Observation of Lattice Defects in Transparent Sheets of Silicon"
   H. J. Queisser and R. H. Finch
   Bull. Am. Phys. Soc. 7, 211 (1962)

3. "Photoresponse of Small Angle Grain Boundaries in Silicon"
   W. W. Hooper and H. J. Queisser
   Bull. Am. Phys. Soc. 7, 211 (1962)

4. "Growth of Silicon Bicrystals with the Dash Pedestal Method"
   R. Gereth
   J. Electrochem. Soc. 109, 1068 (1962)

5. "Structure and Origin of Stacking Faults in Epitaxial Silicon"
   R. H. Finch, H. J. Queisser, J. Washburn and G. Thomas
   J. Appl. Phys., February 1963

6. "Properties of Twin Boundaries in Silicon"
   H. J. Queisser
   J. Electrochem. Soc. 110, 52, (1963)

7. "Thermal Instability in Power Transistors"
   R. M. Scarlett and W. Shockley
   IRE Solid State Device Res. Conf., Durham, N. H. 1962

8. "Oxygen Impurity Atmospheres Around Dislocations in Silicon"
   H. J. Queisser
   Internat. Conf. on Crystal Lattice Defects, Kyoto (Japan) 1962
   Proceedings to be publ. in J. Phys. Soc. Japan, (1963)

9. "Thermal Instability and Hot Spots in Power Transistors"
   R. M. Scarlett, W. Shockley, and R. Haitz
   Physics of Failure Symposium, Chicago, Sept. 1962 (to be publ.)

10. "Dislocations and Semiconductor Device Failure"
H. J. Queisser
Physics of Failure Symposium, Chicago, Sept. 1962 (to be publ.)

11. "The Importance of Respect for the Scientific Nature of Practical Problems"
W. Shockley
Address at Physics of Failure Symposium, Chicago, Sept. 1962

12. "Microplasmas at Stacking Faults in Epitaxial Silicon"
H. J. Queisser and A. Goetzberger
Bull. Am. Phys. Soc. 7, 602, (1962)

Index

Part I

LATTICE DEFECTS IN SILICON

Part II

THERMAL INSTABILITY IN EXTENDED TRANSISTOR AND THERMISTOR STRUCTURES

Part I

LATTICE DEFECTS IN SILICON

1. INTRODUCTION

The work described in this report is part of the "Physics of Failure" program.[1] The objective of this contract was to investigate failure mechanisms in silicon devices. The scope of the program is limited to bulk effects. The influence of lattice defects upon performance and reliability of silicon devices was investigated, with particular emphasis on dislocations. These results are described in the first part of this report.

It is known[2] that impurity diffusion is enhanced along and parallel to edge dislocations. Such diffusion enhancement can lead to non-uniform doping of devices and thus constitute a failure mechanism. Dislocations can also serve as nucleation centers for the formation of precipitates of foreign atoms. A detailed review of the previous studies of these effects of dislocations on semiconductor properties will not be given here. For an introduction to these problems, the reader is referred to the appendix entitled "Dislocations and Semiconductor Device Failure."

---

1 J. Brauer, "Physics of Failure" RAS-TM-62-2, May 1962, Applied Res. Lab., Rome Air Development Center.

2 H. J. Queisser, K. Hubner, W. Shockley, Phys. Rev. 123, 1245 (1961)

A particularly favourable way to study effects of dislocations consists in utilizing bicrystals with small angle grain boundaries. When two crystal grains are slightly misoriented with respect to each other, the boundary between these two grains is defined by an array of regularly spaced edge dislocations. The effects of the individual dislocations are superimposed and can be easily observed. Methods for the growth of bicrystals have been described in the literature[3]; an improvement of the technique[4] will be presented later in this report. Theory and experimental methods of diffusion measurements along small angle grain boundaries in silicon have also been reported in detail earlier.[2]

Further measurements on impurity diffusion and precipitation phenomena are described in this report. Next, photoelectric investigations are discussed which suggest that impurities strongly influence the electrical behavior of edge dislocations in silicon. This result indicates the need for still cleaner and better defined grain boundary material. Consequently, a method was devised to grow "oxygen-free" bicrystals. This work was performed under partial sponsorship of this contract, but is reported in full here.

---

3 H. F. Matare and H. A. R. Wegener, Z. Physik, 148, 631 (1957)

4 R. Gereth, J. Electrochem. Soc., 109, 1068 (1962)

The following portions of the report deal with other lattice defects. Twin boundaries, both coherent and incoherent, are investigated, in particular with respect to diffusion properties and influence on p-n junction devices. The final sections describe a technique of producing very thin sheets of silicon and the observations of defects in these thin sheets. Stacking faults and partial dislocations have been detected in epitaxial silicon layers by this method. The origin and structure of these faults are discussed in detail. These defects have a direct influence on device characteristics since preferential microplasma breakdown has been observed at the stacking faults.

2. GRAIN BOUNDARY DISLOCATIONS

This section covers the investigations on small angle grain boundaries in silicon. The section consists of three parts. In the first part (2A), we report on impurity diffusion and precipitation effects. Photoresponse measurements are presented in part 2B. The third part (2C) describes the growth of bicrystals with the "pedestal method."

A. Impurity Diffusion and Precipitation

(1) Diffusion Measurements

Previous studies[2] have shown that donors and acceptors diffuse much more rapidly along small angle grain boundaries than in the undisturbed bulk silicon. This effect was explained[2] on the basis of the individual edge dislocations which form the boundary. Edge dislocations

attract foreign atoms;[5] therefore impurity "atmospheres" are created around the dislocations. These "Cottrell atmospheres" constitute a concentration enhancement which is thought[2] to be one reason for the more rapid impurity diffusion along the dislocations. A possible increase in vacancy concentration and a possible decrease in the activation energy for the diffusion jump may also contribute to the enhancement. Two coefficients are used in our analysis[2] to compare the grain boundary dislocation to the undisturbed material. W_o is the equivalent width of a slab of "good" material which contains as many impurities as does the grain boundary. W_D is the width of a slab which carries the same diffusion current as does the grain boundary. (The exact mathematical formulation is given in Ref. 2). These two coefficients describe the enhancement of diffusion at the dislocations. It has been shown[2] that in principle both W_o and W_D can be measured by observing the spike-shaped diffusion front of a p-n junction at a grain boundary. The angle 2θ of the spike at the tip and the velocity of advance v of the tip can be correlated. The following relation between those two measurable quantities and the coefficients holds:

$$(v \tan \theta)^{-1} = (W_D \sin^2 \theta - W_o)/2D \qquad (2.1)$$

The previous measurements have shown that this relation is fairly well obeyed for phosphorus diffusing down the grain boundary of

---

[5] A. H. Cottrell, Dislocations and Plastic Flow in Crystals, Clarendon, Oxford (1953)

a p-type bicrystal. Other dopants such as boron, antimony, indium, gallium, bismuth, have been investigated in some preliminary experiments[6]. It was observed that the values for the diffusion enhancement increased with increasing misfit of the doping atom in the silicon host lattice. This is to be expected from the predictions concerning the Cottrell attraction of impurities to dislocations. However, the results obtained for the acceptors boron and gallium had indicated some discrepancies and were not as reproducible as the results for the donors. It was therefore decided to study these effects further under the present contract.

The practical aspects of the enhanced diffusion are obvious. These effects can lead to considerable doping non-uniformities which cause malfunctioning or device failure. For example, short circuits through diffused transistor base layers may result from the enhanced diffusion along dislocations.* Thus, there is a direct consequence for the reliability of silicon devices. Ultimately one should be in a position to calculate and predict the extent of the detrimental effects which are caused by the enhanced diffusion in devices. Such predictions require a complete and quantitative description of the basic phenomenon of grain boundary diffusion. The measurements to be described here represent part of a program to obtain such a quantitative description.

---

[6] H. Queisser in Interim Sci. Rep. 1, Contract AF 19(604)-8060, August 1961

* See Appendix for specific examples, obtained prior to this contract.

Grain boundary diffusion experiments with gallium were carried out in the following manner.

Bicrystals were grown both by the Czochralski method[7] using two different seeds, one for each grain (as described by Matare)[3], and by the pedestal method (as described in Section I-2C of this report) using a single seed, containing a grain boundary in the desired orientation and of predetermined angle of misfit. The bicrystals were grown with a [100] pull axis and grain boundaries in either (100) or (110) median orientation. The angles of misfit between grains ranged between 3° and 10.5°. The misfit is determined by the seed holder and can be checked with standard X-ray methods or by making use of diffusion-induced slip patterns on the surface of a slice containing the grain boundary[8].

The latter method was chosen for its ease of application. Samples were obtained by slicing the crystals perpendicular to the growth axis. The resulting slices were mechanically lapped and polished. Before the samples were used, they were subjected to chemical cleaning steps to remove surface damage and surface contamination.

---

[7] J. Czochralski, Z. Physik Chemie., 92, 219 (1917)

[8] H. J. Queisser, Appl. Phys. 32, 1776 (1961)

a. Diffusions

The diffusions were carried out in an open tube furnace with a simple hot zone, in which the temperature could be held constant within $1^{o}C$. Doping was obtained from the vapor phase through two different methods: either doping with Ga-doped silicon powder or from Ga_2O_3 with graphite. Of these, only the method based on Ga_2O_3 and graphite proved to be reliable and reproducible. Therefore, this method was used throughout the experiments. Equal parts by weight of Ga_2O_3 and graphite powder were thoroughly mixed and placed at one end of a quartz boat. Heavy surface pitting will occur unless the slice is protected by oxidation. Therefore, the samples were preoxidized, then placed approximately one inch from the source in the same boat. The boat, fitted with a loose cover, was then placed in the hot zone of the furnace, with dry nitrogen as a carrier gas.

b. Observations

Observations of the diffusion profile were made with the use of beveling and staining techniques. Of several known staining techniques, the one used by Fuller and Ditzenberger[9] was found the most reproducible.

This method utilizes a solution of 0.05% conc. HNO_3 in a 50% HF, 50% H_2O solution. A drop of this solution, placed on a beveled surface containing a p-n junction, will produce junctiondelineation under strong illumination of the surface. The p-type region will be dark colored

---

[9] Fuller, C. S. and Ditzenberger, J. A., J. Appl. Phys., 27, 550 (1957)

with respect to the unstained n-type region. In mounting the sample on the beveled slug to facilitate beveling of the sample, deviations of the predetermined angle of lap can occur due to irregularities in the sample-slug interface. Determination of the exact angle of lapping then becomes necessary. This was accomplished through interference microscopy. In this technique fringe patterns are obtained which make accurate determination of the junction depth possible, independent of the angle of lap. Using photographs of known magnification of the entire beveled surface, the exact angle of lap is easily found and the spike depth and spike tip angle can then be reconstructed.

During the course of the experiments it was found that the length of the stained spike was sensitive to staining time. It therefore became necessary to establish some standard technique to insure reproducible and comparable results. Fig. 2. 1 shows the apparent spike length as a function of the time of staining. A saturation value seems to be reached after an initial increase of the apparent length. After this phenomenon was discovered, the "saturation" value of y_s was assumed to be the correct one. There is, however no assurance that this is really the case, as will be pointed out later on. Repeated measurements indicated that no increase of junction depth y_j off the boundary could be observed after prolonged staining.

The reconstruction of the spike tip posed considerable difficulties. Photographs were taken at high magnifications (up to

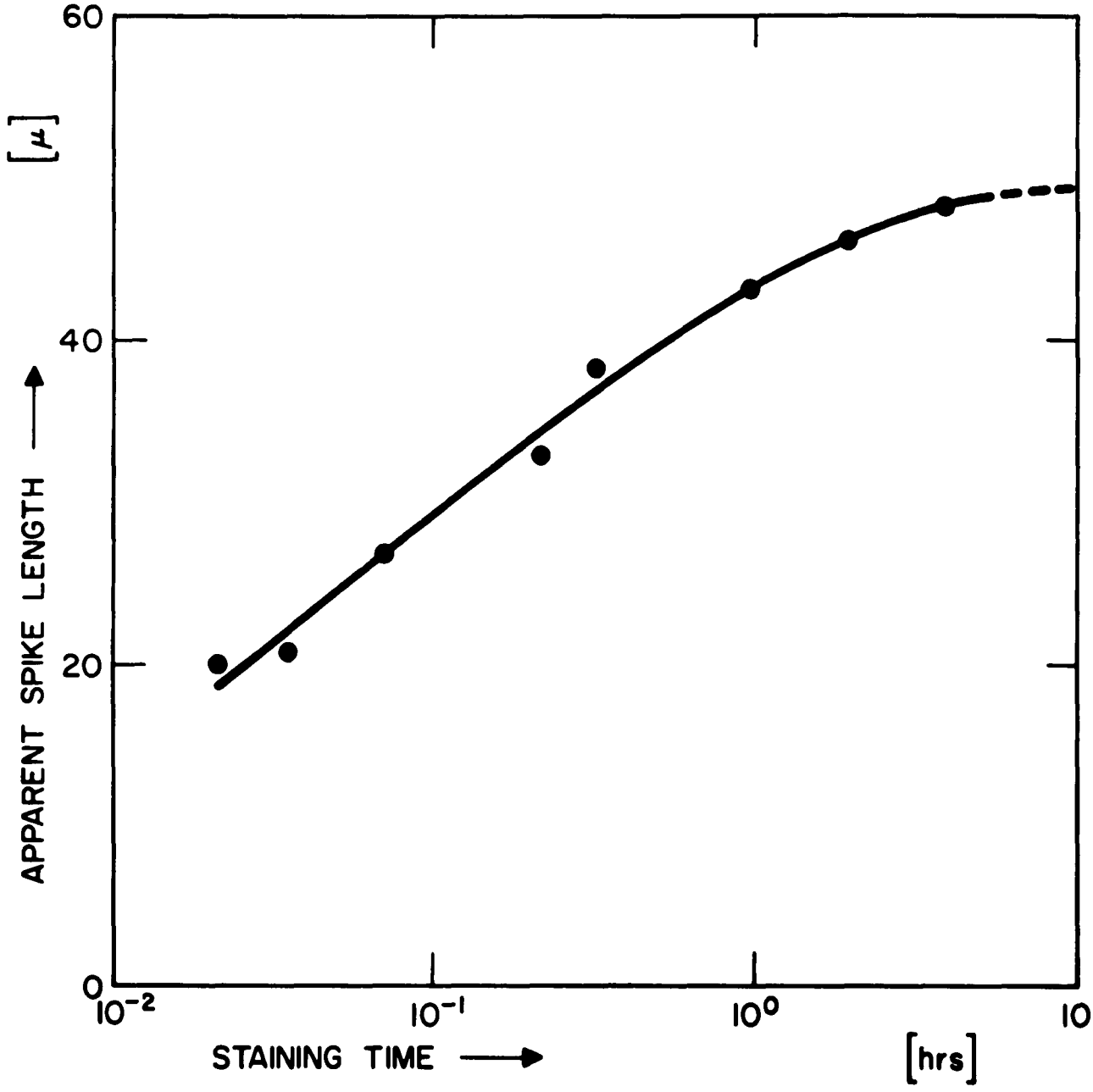


Figure 2.1

Apparent spike length as a function of staining time

1000X); from these photographs the profiles were reconstructed point by point. The definition of the p-n junction was often insufficient, especially at the spike tip. In a number of cases the angle θ could only be extrapolated through straight line approximations from the spike flanks.

c. Experiments

Diffusion experiments were carried out, under constant surface concentration conditions as well as under constant impurity charge conditions. The temperatures at which these diffusions were carried out were 1000 C, 1100°C and 1200°C. Surface concentrations at 1000°C and 1100°C were too low to be useful. Therefore, only the results obtained with diffusions performed at 1200°C will be considered. Several slices each from the following bicrystals were used:

1. 3° angle of misfit; (110) median orientation of grain boundary; 0.2 Ω cm n-type.
2. 6° angle of misfit; (100) median orientation of grain boundary; 0.4Ω cm n-type.
3. 3° angle of misfit; (110) median orientation of grain boundary; 25Ω cm n-type.
4. 10° angle of misfit; (100) median grain boundary plane 0.4Ω cm n-type

The experiments were made up to a maximum of 10 hours diffusion time, with data taken at 1 hour intervals, whenever possible. The following data were obtained: junction depth y_j and sheet conductance $\bar{\sigma}$ of the undisturbed material, spike depth y_s and spike angle 2θ as a function of diffusion time t. From y_j and $\bar{\sigma}$ the regular diffusion coefficient D

of gallium in the undisturbed silicon can be calculated; the values thus obtained were within at least 10% of the literature value.[9] The evaluations of the grain boundary measurements are described next.

d. Results and Discussion

An attempt was made to plot the results of these experiments so as to test the validity of the "spike-velocity equation" (2.1). This is done by plotting $(v \tan \theta)^{-1}$ versus $\sin^2 \theta$. The velocity v can be obtained by a graphical differentiation of the spike depth y_s as a function of time t. It is seen that the half-angle θ of the spike tip enters twice into Eq. (2.1), namely as $(\tan \theta)^{-1}$ and as $\sin^2 \theta$. Therefore, the angle θ is the most critical quantity involved. When this angle is reasonably large it can be determined with sufficient accuracy; this is the case for impurities with relatively small enhancement factors for grain boundary diffusion, such as phosphorus. The gallium spikes, however, are much longer and narrower with small tip angles. The measurement accuracy therefore, is very low. The gallium data obtained were in most cases not accurate enough to obtain unambiguous values for W_D and W_o with the spike velocity method.[2] Some of the latest series of measurements, however, gave a reasonably close fit of Eq. (2.1). Straight lines resulted in plots of $(v \tan \theta)^{-1}$ versus $\sin^2 \theta$. These measurements were done very carefully with a standardized staining procedure. Thus, they may be regarded with more confidence; the results may be considered to yield at least the order of magnitude for the parameters W_o and W_D. These

data are presented in condensed form by the following table.

Table 2.1

Gallium Diffusion Runs on (100) Grain Boundary, 10.5° Misfit

Crystal No.	Doping	Diff. Temp.	W_o	W_D
2569	P 0.4 Ω cm	1200°C	0.6 μ	290 μ
2158	P 0.4 Ω cm	1200°C	0.1 μ	160 μ

Evaluation with the "spike-velocity" method (Ref. 2 and 6)

These values do not appear unreasonable. W_o can be estimated by theory (See Eq. 10 of Ref. 2) from the misfit of the gallium atom in the silicon lattice. Such an estimate gives W_o = 1 micron, which is not too different from the experimental results. As mentioned before, the coefficients presented here can be regarded only as yielding the order of magnitude, since thus far the experimental conditions are not well enough under control. We shall next examine this question of the inadequate experimental methods in more detail.

The most critical technique involved in these measurements is undoubtedly the staining. Considerable attention was given to this problem, with the hope of eventually developing this method to the high degree of reproducibility which is required for the exact delineation of long p-type diffusion spikes. The most important result, which hasbeen mentioned already, is the dependence of the apparent spike length upon staining time. This influence of the duration of the staining process partly explains why

consecutive lapping and staining of the profile had given nonreproducible results, while simultaneous checks of the junction depth indicated that not merely the doping method was unreliable.

Occasionally spike delineations were obtained which were unexpected and difficult to explain on the basis of a diffusion process. For example, "spikes" were observed to display no tip, but a blunt end. The staining probably yields a deceptive result here. It became rather doubtful whether the staining method in all cases truly indicates the diffusion front or rather some inherent property of the boundary itself. This question was further examined. A high resistivity n-type bicrystal was beveled and then subjected to a long staining procedure. It was observed that the grain boundary stained dark (p-type) through its entire length. The grain boundary itself behaves under these conditions as if it were a p-type sheet. Thus, it is not surprising to often find wide scatter in the data for the spike diffusion depth y_s.

These findings in turn prompted photoresponse measurements which will be described in detail below. The main results of these photoresponse studies indicate that in silicon a small angle grain boundary can appear to be either n- or p-type, depending upon the thermal history of the crystal. The staining agrees with the photoeffect data. The assumption of impurity atmospheres around dislocations seems to explain the results best. In particular, oxygen seems to be influential, as will be shown below. Oxygen is present in crucible-grown crystals

in large quantities and it is known[10] to be unevenly distributed. This fact could explain the observations that the spike length was not always uniform over a diffused slice. The staining might give erroneous results because of the variations of the impurity atmospheres but also the gallium diffusion itself may be altered by presence or absence of uncontrolled impurities. These impurities can occupy sites at the dislocation core, thereby changing the strain field and possibly also the elementary diffusion jump.

Fast diffusing acceptors seem to be the dopants most difficult to measure accurately at grain boundaries, for several reasons. First, the rapid diffusion causes narrow spikes with small tip angles, which are difficult to measure. Secondly, the fact that the boundary stains dark (p-type) after a heat treatment, means that the spike end cannot be determined with certainty. Furthermore, it may be that oxygen is present around the boundary dislocations. Oxygen and Group III-acceptors can undergo complicated solid state-reactions[11] which might also be influential. Finally, the acceptor Al is known to have a great tendency to form metallic precipitates at dislocations[12]. Possibly this tendency is also prevailing for Ga, which might influence grain boundary diffusion.

---

10 G. H. Schwuttke, J. Electrochem. Soc. 108, 163 (1961)

11 C. S. Fuller et al., J. Phys. Chem. Solids 13, 187 (1960)

12 R. Bullough et al., J. Appl. Phys. 31, 707 (1960)

These results led to the conclusion that the "spike-velocity method" was unsatisfactory for the presently available grain boundary material, at least for gallium and probably also boron, which had previously been found to exhibit anomalies. The emphasis of the research was therefore shifted towards the three following objectives:

Measurement of grain boundary diffusion with techniques which do not rely as strongly on staining.

Further study of the effects supposedly resulting from atmospheres of uncontrolled impurities at the dislocations.

Preparation of clean, well-defined silicon bicrystals.

Progress was made along the lines of all three objectives. Oxygen-free bicrystals could be grown with the pedestal method (see section I-2C). The evaluation of this material could not yet be started on a large scale. The photoresponse data (section I-2B) gave considerable information about the influence of impurity atmospheres on the electrical behavior of the dislocations. Finally, conductivity measurement techniques were developed, which promise to become a useful alternate method for measuring grain boundary diffusion in semiconductor crystals. These methods will be described next.

(2) Conductivity of Diffusion Spikes

This portion of the report describes electrical conductance measurements made on grain boundary diffusion spikes. As mentioned earlier, this technique was initiated as a means of evaluating grain boundary diffusions while avoiding, where possible, the difficulties encountered in staining.

Three basic structures were used for these measurements, shown in Figs. 2.2a,b, and 2.3. Fig. 2.2a shows the spike profile resulting from a gallium grain boundary diffusion. The spike conductance is measured between the top and bottom surface of the sample connecting the spike. The sample thickness is approximately 150 μ. It can be seen that in order to obtain a connection via the diffusion spike for this thickness, an impurity such as gallium must be chosen, which diffuses down the grain boundary very rapidly. For slower diffusants, thinner samples could. of course be used, but they become difficult to handle.

Epitaxially grown layers provide an excellent solution to the problem of slice thickness and diffusant requirements just outlined. Fig. 2.2b shows the second structure used; an epitaxial layer (dark stain) grown onto a bicrystal slice. The grain boundary is propagated through the epitaxial layer. A subsequent diffusion into the epitaxial layer produces the spike profile shown. The diffusion spike now connects the substrate with the diffused layer. This configuration eliminates the restrictions on allowed diffusants while retaining the thick slices for ease of handling. This geometry has one disadvantage, however. There is a

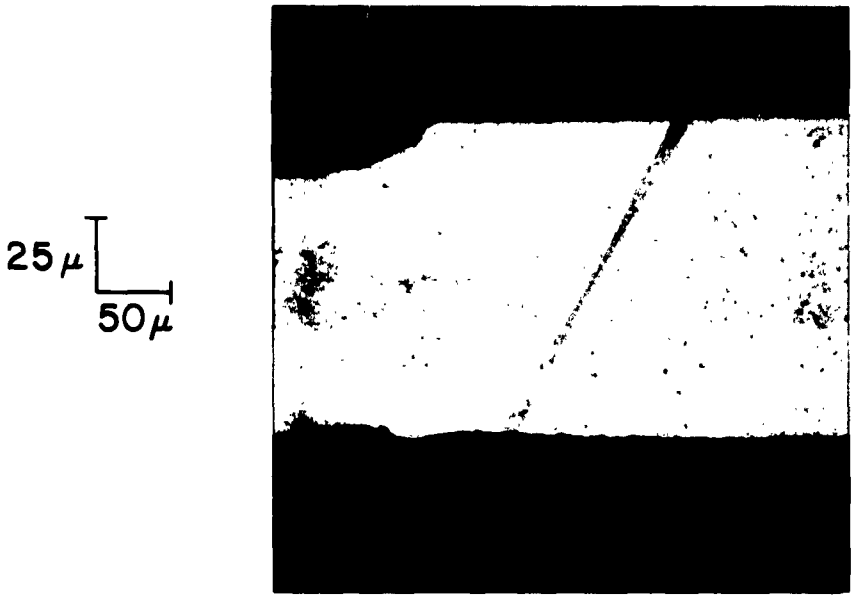


Figure 2.2a

Gallium Grain Boundary
"Spikes" through a 125 μ slice

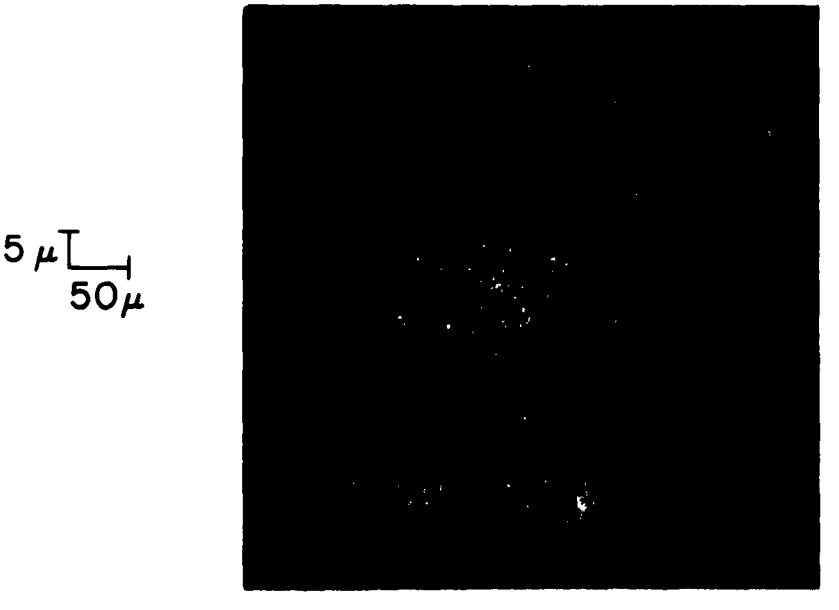


Figure 2.2b

Phosphorus Grain Boundary
"Spike" through an Epitaxial
Layer

spreading resistance in the regions where the spike connects to the surface diffused layer and the substrate. Since the diffused layer is very thin and degenerate, the spreading resistance at this point is negligible. This is not true, however, at the point where the spike connects to the substrate. Since this spreading resistance is in series with the total spike resistance its effects can only be minimized by making the spike resistance large compared to the spreading resistance. This is best accomplished by using long narrow diffusion spikes.

Fig. 2.3 shows the third structure used for grain boundary conductance experiments. The conductance is measured along the grain boundary spike. Contacts are made through isolated mesas on the diffused surface of the bicrystal slice. Small circular areas over the boundary are masked with wax while the rest of the diffused layer is etched off. Two adjacent measas are electrically connected through the diffused wedge only. This configuration has distinct advantages over the two previously mentioned structures. Contacting is done from one side only, as this simplifies the experiments. The rather critical condition of two meeting spikes is avoided. The magnitude of the conductance can be controlled by the choice of contact spacing. Finally, the geometry lends itself well to measurements of spike conductance versus stepwise removal of the top portions of the wedge. This provides information about the impurity distribution in the spike. The results of these investigations will be described later.

The sample preparation for the "meeting spike" structure of Fig. 2.2a was carried out in the following manner. Bicrystal slices of

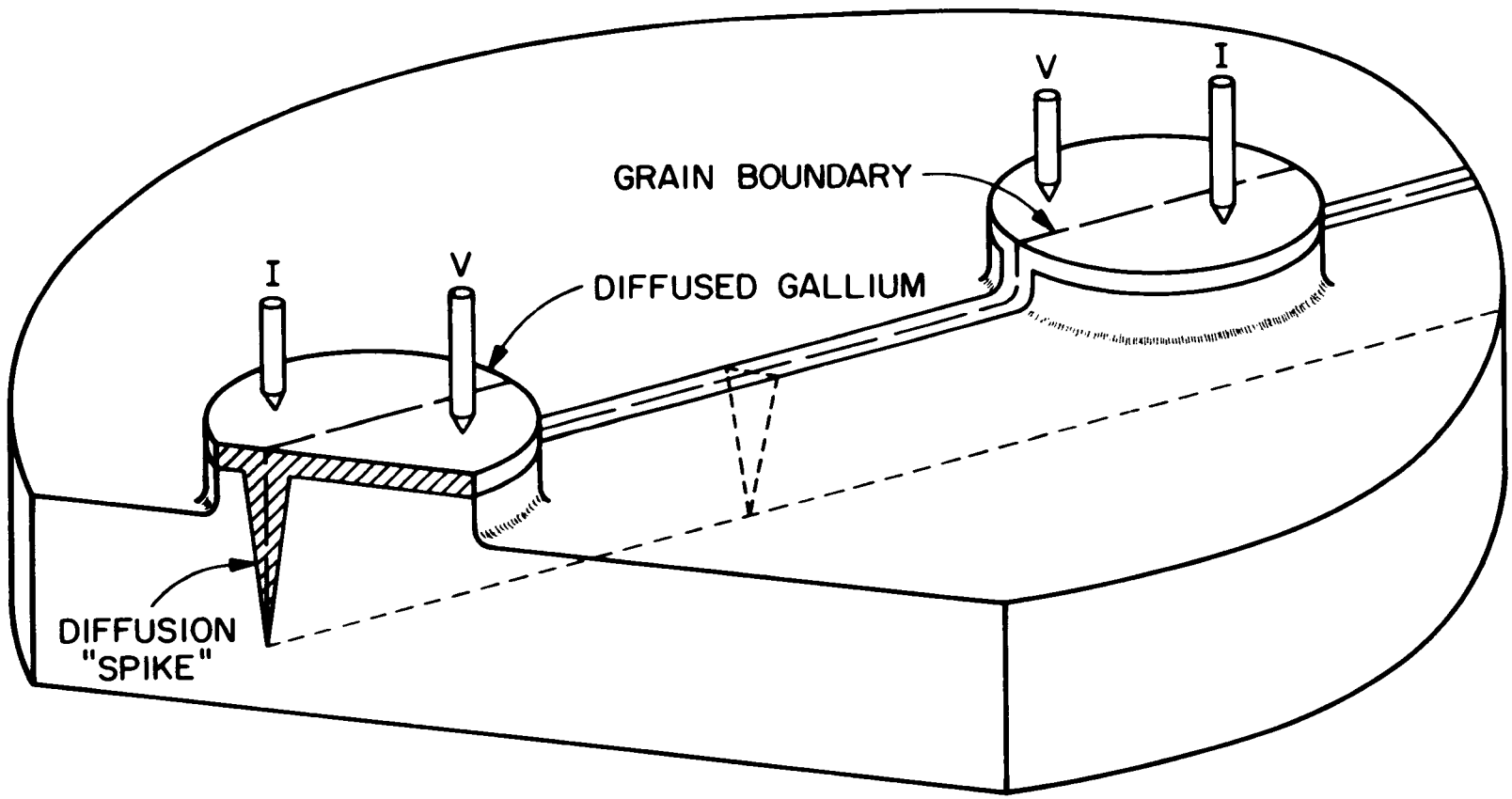


Figure 2.3

Structure used for measurement of
diffusion spike conductance vs etching depth

7-12-62
50285

0.4 ohm-cm n-type resistivity, 150 μ thick, were used for a gallium diffusion. Diffusion was done by predeposit from a Ga_2O_3 source at 1200°C, with diffusion time varying with different samples from one half-hour to four hours. The diffusion takes place from both sides of the slice, with the resulting profile shown in Fig. 2. 2a. The one-half hour diffusion yields spikes of low conductivity, measured between the diffused layers. The four hour diffusion gives spikes of high conductivity. Following the gallium diffusion, isolated devices were made by masking 1.5 mm square areas lying on the grain boundary, and etching away the unwanted material. The grain boundary on the top surface of the device was masked along its 1.5 mm length, approximately 0.2 mm wide, and the unmasked gallium surface removed by etching. The resulting structure is a 1.5 mm square, with a diffused gallium layer on the bottom surface, and a gallium "ridge" 0.2 mm x 1.5 mm on the top surface. These two gallium layers are connected by the grain boundary diffusion spikes.

Electrical contact was made to these samples using four isolated probes; two each for current and potential measurements. Contact to the top gallium "ridge" was made by placing two small probes on this surface. Two isolated contacts were made to the bottom gallium layer by placing the device on a glass slide with two evaporated gold strips on its surface. The device was placed in such a manner that the strips are parallel to the grain boundary, one on each side, but not contacting the boundary. The current was

applied between one top and bottom contact, and the potential measured between the remaining two contacts. Current-voltage measurements were made over the current range of 10^{-6} A to 10^{-3} A, and voltages were measured with a potentiometer. In all cases where lapping and staining showed meeting diffusion spikes, the current-voltage plots were ohmic. The range of spike conductance was 7×10^{-5} mho for a sample diffused for one half hour to 1×10^{-2} mho for samples diffused four hours. These figures represent the extreme measured values, rather than averages. It was found while making these measurements that the grain boundary spike conductance did not vary with diffusion time as expected. Two phenomena were observed. First, for a given diffusion time two devices adjacent to each other on the same slice showed considerable difference in their spike conductances, sometimes as large as a factor of 10. Second, the rate of change of spike conductance vs diffusion time was not the same for adjacent devices, and in many cases the conductance first increased and then decreased with additional diffusion. The major effects of the conductance occur at the "waist" of the spikes where the tips overlap. This is the narrowest region; also the impurity concentration of the spikes is lowest there. Because of this, conductance measurements on these "narrow-waisted" spikes are most sensitive to grain boundary diffusion anomalies. The conductance of the "meeting spike" geometry is described according to the diagram of Fig. 2.4. The solid lines represent surfaces of isoconcentration, changing by a factor e^{-1} between adjacent lines. $v_x = v \sin \theta$ represents the

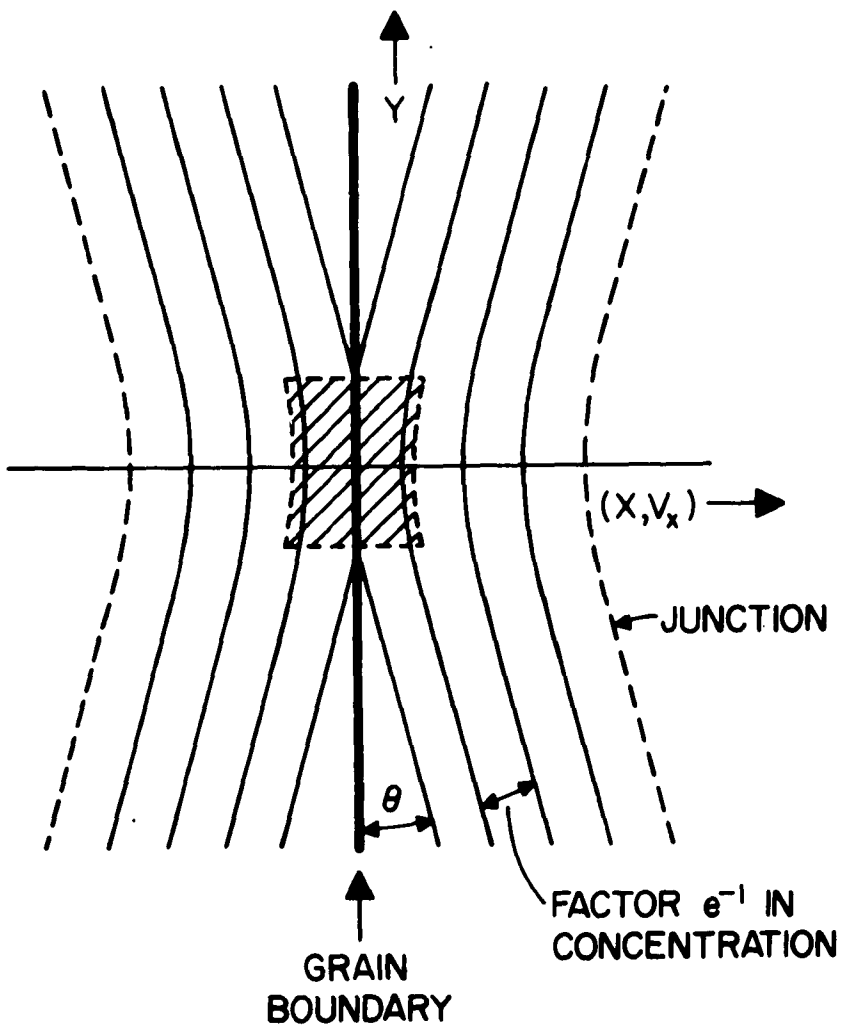


Figure 2.4

Schematic representation of meeting diffusion spikes. Isoconcentration lines are indicated; the center section yields the major portion of the electrical resistance.

velocity of the diffusion front at the spike waist, in the x direction, (v is again the spike velocity $v = dy_s/dt$ (see Page 11) and θ is the spike half angle[2]. The build-up of conductance should increase exponentially for small diffusion times after the spikes first meet. If the variation in conductance is exponential with time, then

$$d \ln G/dt = \gamma \tag{2.2}$$

where γ is a constant, and G the conductance. An expression for γ in terms of the parameters θ, v, and the diffusion constant, D, may be found as follows. Let c = the concentration at the waist of the spike.

$$c = c_o \exp - a\,[x \cos\theta + (y - vt) \sin\theta] \tag{2.3}$$

where c_o = the concentration at the boundary. Substituting u for $(x \cos\theta + y \sin\theta)$,

$$c = c_o \exp - a\,[u - vt \sin\theta] \tag{2.4}$$

and

$$\partial^2 c/\partial u^2 = a^2 c = \partial c/D\partial t = av \sin\theta\, c/D \tag{2.5}$$

$$\partial^2 c/\partial u^2 = v \sin\theta\, \partial c/D\partial u \tag{2.6}$$

and

$$a = v \sin\theta/D \tag{2.7}$$

so that

$$c = c_o \exp(-v \sin\theta\,[x \cos\theta + (y - vt) \sin\theta]/D) \tag{2.8}$$

since

$$\partial c/dt = v^2 \sin^2 \theta \, c/D \qquad (2.9)$$

the build-up of conductance is thus proportional to exp $(v^2 \sin^2 \theta t/D)$. Therefore

$$\gamma = d\ln G/dt = v^2 \sin^2 \theta/D = v_x^2/D \qquad (2.10)$$

and

$$G = G_o \exp (v_x^2 t/D) \qquad (2.11)$$

where G_o = conductance at diffusion time t_o when the spikes first meet, and G is the conductance after additional diffusion time t.

Measurements of the "meeting spike" conductance vs diffusion time were made with gallium spikes. The diffusion was done by predeposit from a Ga_2O_3 source for one half hour at 1200°C. Devices were made by etching, as described earlier. Those devices which showed the highest ohmic resistance through the spike were used for conductance measurements. The results of these measurements are shown in Fig. 2.5. The conductance of the grain boundary spike is plotted vs the diffusion time. The shortest time shown is one half hour, which is that of the initial predeposit. Data are shown for two devices taken from the same slice. Diffusions following the one half hour predeposit were done at 1200°C for 15 minute intervals without gallium source. For diffusion times up to one hour following the initial predeposit, the plot of log G vs diffusion time is a reasonably straight line. Longer diffusions cause a deviation from exponential

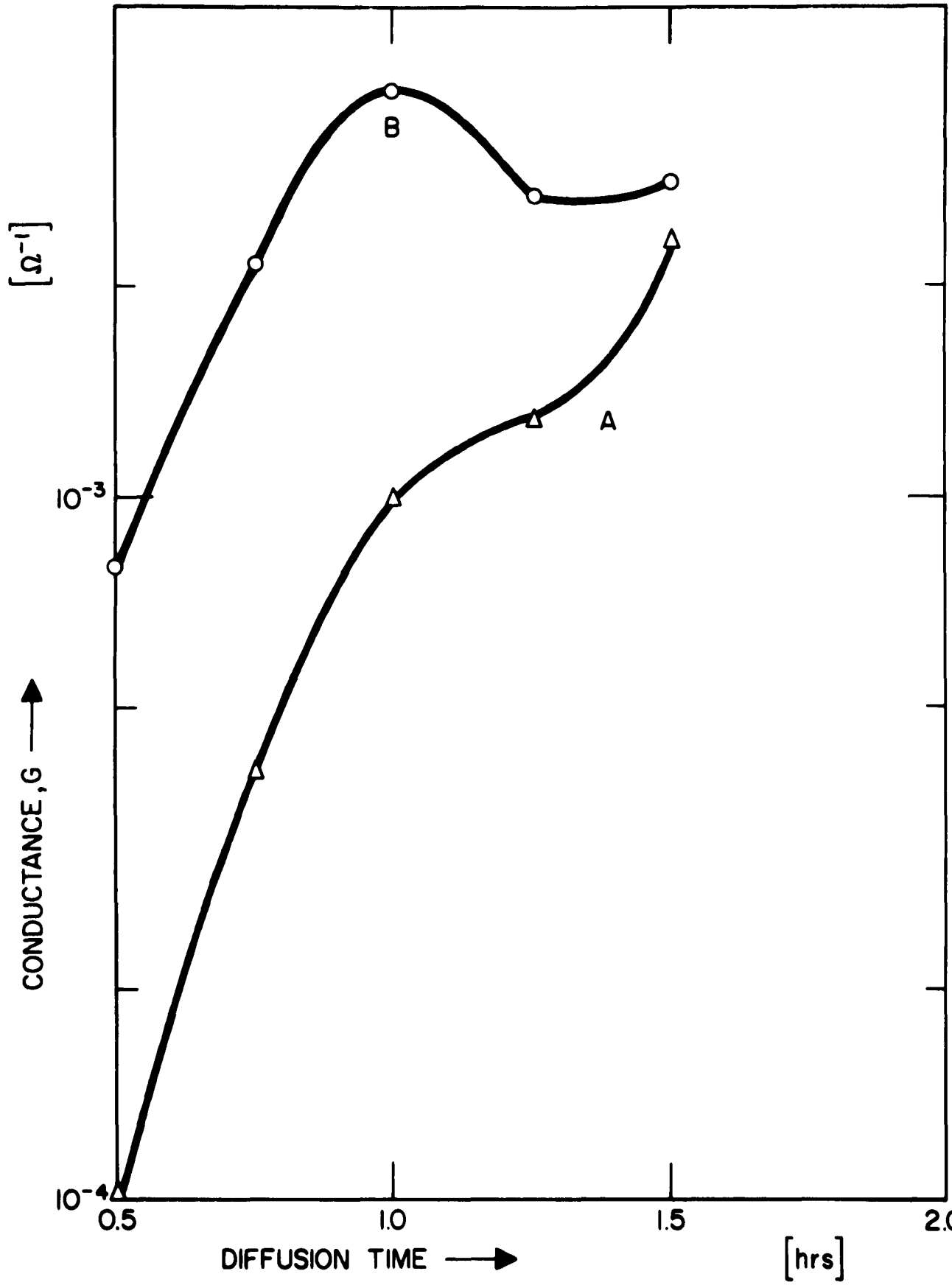


Figure 2.5

Conductance of meeting spikes versus diffusion time

10-17-62
50404

behavior. This could be explained by a loss of dopant in the grain boundary due to out diffusion which is known to be present with gallium. It can be seen from Fig. 2.5 that, although the conductivity of the two samples differs by a factor of 8 following the initial pre-deposit, the initial slopes, γ, are nearly the same for short diffusion times. A comparison of γ determined by electrical measurements could be made with v_x^2/D by means of Eq. (2.10). The determination of this quantity, however, involves lapping and staining after each diffusion interval. The difficulty in obtaining reliable staining results with gallium spikes, together with the unavoidable loss of grain boundary material through lapping renders this comparison of little use. One can, however, obtain a value for W_D making use of these conductivity measurements.

From the spike velocity equation, Eq. (2.1) neglecting in it the small constant W_o, one has

$$W_D = 2D/v \tan \theta \sin^2 \theta \qquad (2.12)$$

Solving Eq. (2.10) for v and substituting into Eq. (2.12) gives for small θ, where tan θ can be considered equal to sin θ,

$$W_D = 2(D/\gamma)^{1/2}/\sin^2 \theta \qquad (2.13)$$

Although determination of the spike angle θ must be made by staining, this can be done quite accurately by extrapolation of the spike profile to the spike tip. Staining errors entering in the estimate of W_D

from Eq. (2.13) are minimized because they enter only in the measurement of θ, whereas they enter twice in Eq. (2.12) through measurement of v and θ. Data for samples A and B of Fig. 2.5 gives $\gamma_A = 5.5$, $\gamma_B = 4.1$, and $\sin\theta = 0.08$. $D = 1.5\ \mu^2/\text{hr.}$[10] From these data, $W_{DA} = 162\ \mu$, and $W_{DB} = 190\ \mu$. These values are in agreement with those of Table I found by diffusion and the spike velocity method. It should be remarked that the samples used for these conductivity measurements were diffused from the same bicrystal, No. 2158, used for determination of W_D presented in Table 2.1. This is important, since we find considerable variation in conductivity measurements and staining results for different bicrystals having the same diffusion treatment. This is further evidence that the gallium diffusions are not well enough defined, and that uncontrolled impurity atmospheres surrounding the grain boundary may have large effects on the grain boundary diffusion.

The second grain boundary spike configuration used for conducitivity measurements makes use of epitaxially grown layers, briefly described earlier with the aid of Fig. 2.2b. Although this structure permits using diffusants of low enhancement such as phosphorus, the estimated resistance of the diffusion spikes is much lower than the spreading resistance. This makes determination of the spike resistance rather difficult. Fig. 2.2b shows a phosphorus spike which illustrates the use of an epitaxial layer. Conductivity measurements utilizing epitaxial layers were carried out with gallium. Very narrow,

long spikes can be easily obtained, with the result that the total spike resistance is much larger than the spreading resistance. A sample was prepared by growing a 20 μ n-type 1 ohm-cm layer epitaxially onto a 150 μ 0.4 ohm-cm p-type bicrystal slice. A subsequent gallium predeposit was done at 1200°C for 25 minutes. The resulting grain boundary spike was 18 μ long and of 5 μ average width. An estimation of the spreading resistance for this spike gives $R_s \sim 3$ ohms. Current-voltage measurements of the spike lie in the range of 10 to 15 ohms, which is considerably larger than the spreading resistance, and therefore a good measure of the spike resistance. In order to determine if these values are reasonable from the standpoint of available diffusion data for gallium, spike resistance limits were calculated based on two conditions of impurity concentration at the grain boundary during a total diffusion time t_D. The first estimate gives a lower limit of the spike resistance. We assume that the impurity concentration everywhere at the boundary within the spike is $C_B = 0$ for times $t < t_B$, and $C_B = C_S$ (surface concentration) for $\geq t_B$. The time t_B can be obtained by assuming that $t_D - t_B$ is the diffusion time required to produce the observed maximum spike half width. With this assumption and the aid of diffusion data, the lower limit of the spike resistance for these samples is 3 ohms. This value may be considered a lower limit since the assumed boundary concentration $C_B = C_s$ is much too high.

The second estimate gives an upper limit to the spike resistance. The condition is that the boundary concentration is constant in

position and time along the boundary, and the concentration is that required to produce the minimum spike half width observed. This assumption, together with diffusion data, gives an estimate of 17 ohms for the upper limit of spike resistance. This may be considered to be a maximum value since the boundary impurity concentration assumed for this situation is definitely underestimated.

A critical and accurate analysis of the conductivity of these grain boundary spikes should be made by considering the impurity distribution at every point along the boundary with respect to time t and position y, rather than estimating extreme values between which experimental data should lie. The need for a more accurate estimate of the grain boundary impurity distribution led to the new and useful measurement of transverse conductance, described in the beginning of this section.

Samples for this measurement shown schematically in Fig. 2.3 were prepared using both phosphorus and gallium diffusions. The gallium diffused spike will be discussed first. The spike profile was obtained by a gallium diffusion into a 200 μ thick 0.4 ohm cm n-type bicrystal slice. Diffusion was done for 20 minutes at 1200°C from a Ga_2O_3 source. The resulting diffusion spike was $\sim$ 35 μ deep, with a spike half angle $\theta \sim 5^\circ$, and surface junction depth of 2.5 μ. Mesas were made every 1.3 mm along the boundary by masking and etching away the unwanted 2.5 μ diffused layer. Electrical contact was then made to the surface mesas. The conductivity measurements are now made between the mesa "contacts"

as illustrated by the drawing of Fig. 2.3. For a given diffusion it is now possible to measure the change in spike conductance as small portions of the spike are removed from the surface by controlled etching. This measurement is quite useful in evaluating grain boundary diffusions since it allows determination of the impurity distribution at several points along the boundary between the surface and the spike tip. The data for these measurements are plotted in Fig. 2.6 as the wedge conductance in mhos vs the top layer removed, in microns. The measurement is made along the boundary between two mesa contacts 1.3 mm apart.

The etching depth is determined by use of a standard thickness comparator. It can be seen that the conductance decreases exponentially at first, but should approach zero as the entire spike is etched away. Junction leakage effects, however, became dominant when etching was carried further than 20 μ. Similar measurements have been made with phosphorus spikes. These spikes are $\sim$ 15 μ deep; quite shallow compared to gallium, and controlled etching is therefore very difficult. A single etch of 5-10 μ already produces a very high resistance spike, and junction leakage current becomes troublesome. A single conductance measurement of the phosphorus spike following removal of the surface layer, however, is quite useful, and will be discussed later. The plot of Fig. 2.6 gives a measure of the impurity distribution profile along the boundary. The impurity concentration limits at the surface and spike tip are known from four point probe resistivity measurements.

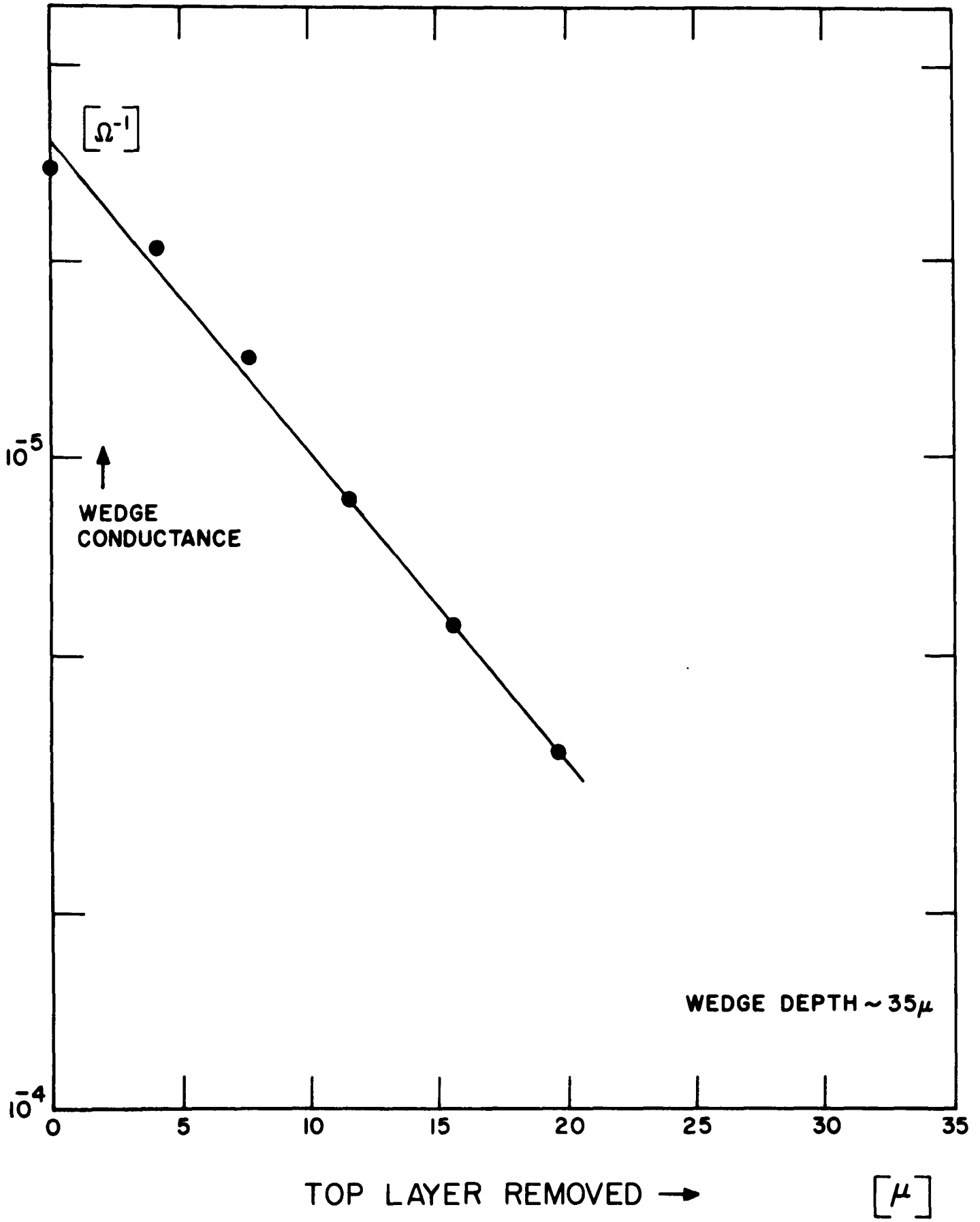


Figure 2.6

Wedge conductance vs top layer removed for gallium-diffused spike

9-25-62
50344

The concentration at the spike tip is the bulk dopant concentration. When the spike length is determined by staining, estimates of the impurity distribution can be made. It now remains to determine how well our assumption of an exponential impurity distribution profile fits the actual grain boundary diffusion conditions. This was done by finding an expression for the average conductance of a thin slab in the diffusion spike, and summing these conductances over the entire spike length. This gave an estimated conductance for the spike which could be compared with electrical measurements, thereby indicating whether or not our assumption concerning the distribution was reasonable. We can derive an expression for the effective conductance $d\bar{G}$ of a thin slab of the spike in the following manner. Referring to Fig. 2.3, the x direction is perpendicular to the plane of the grain boundary, the z direction is measured along the boundary between the mesas, and the y direction is positive upward* perpendicular to the mesa surface, with the spike tip as the origin. Consider a slab dy thick, and z long, located at +y from the spike tip. The width x of the slab is also a function of y because of the wedge-shaped profile. The effective conductance in the z direction is therefore

$$d\bar{G}(y) = \bar{\sigma}(y) \cdot x(y) \cdot dy/z \qquad (2.14)$$

z is the mesa spacing (1.3 mm for the data of Fig. 2.6) and x(y) is simply $x = 2y \tan \theta$. For a spike half-angle $\theta = 5°$, this becomes $x = 0.12\,y$. We have assumed that the impurity distribution down the

* The convention is to consider y positive downward with the origin at the surface.[2] Making y positive upward with the origin at the spike tip, however, greatly simplifies the calculations.

grain boundary is exponential, and now consider the limits of effective conductivity of these slabs within the spike. The lower limit of spike conductivity is that of the bulk material σ_B, at the spike tip, and the upper limit is that of the effective surface conductivity $\bar{\sigma}_s$. Since the conductivity is proportional to the impurity concentration, the effective conductivity of the slab as a function of y from the spike tip is

$$\bar{\sigma} = \sigma_B \exp (y/A) \tag{2.15}$$

where A is the change in y required for $\bar{\sigma}/\sigma_B$ to change by a factor e. For the spikes measured , A = 30 μ. Substitution of Eq. (2.16), and x = 0.12 y into Eq. (2.15) gives

$$z d\bar{G} = 0.12\ \sigma_B \exp (y/30)\, dy \tag{2.16}$$

Putting in $z = 1300\ \mu$, and $\sigma_B = 2.5 \cdot 10^{-4}$ mho $\cdot\ \mu^{-1}$,

$$d\bar{G} = 2.3 \cdot 10^{-8} \exp (y/30)\ y\, dy \text{ (mhos)} \tag{2.17}$$

Integrating over the entire spike length of 35 μ

$$\bar{G} = 2.3 \cdot 10^{-8} \int_0^{35} \exp (y/30)\, y\, dy$$

$$= 2.07 \cdot 10^{-5} \left[\exp \{y/30\} (y/30-1) \right] \Big|_0^{35} \tag{2.18}$$

and

$$\bar{G} = 2.84 \cdot 10^{-5} \text{mhos} \tag{2.19}$$

This value of estimated spike conductance is within 10% of our average measured value $3.1 \cdot 10^{-5}$ mhos, shown in Fig. 2.6 as the point pertaining to the total wedge conductance before any top layer was removed.

Comparison of estimated spike conductance with the measured spike conductance has been made for the various etching depths, and the agreement between these values is reasonably good. The deviation between measured and estimated values becomes greater the further the etching is carried out because of junction leakage current. Therefore, these measurements do not give enough information about the impurity distribution at the spike tip. The foregoing analysis of the grain boundary impurity distribution should, of course, be carried out for distributions other than that of a Gaussian, or exponential. A rigorous analysis, however, using more complicated functions would necessitate the use of computers.

As mentioned earlier, the phosphorus diffusion spikes were too shallow to use for measurements of conductivity vs stepwise etching. A single measurement of the entire wedge conductance, however, allows one to make a first approximation as to the enhanced diffusion due to dislocations. It has been shown[2] that the grain boundary diffusion current per dislocation is larger than the diffusion current for a column of undisturbed crystal material of area b^2 by the "boost factor" $m = SW_D/b^2$. The equation defines the parameter W_D; S is the spacing of the grain boundary dislocations having a Burgers vector b. Assuming that the total dopant concentrations are proportional to the diffusion currents, there should be m times as much dopant through diffusion along one dislocation as through diffusion in a column of area b^2 in undisturbed dislocation free material. This probably underestimates m, particularly with deep spikes as found with gallium. m is obtained by

comparing the wedge conductivity to that of the diffused layer in the dislocation-free material. Since the dislocation spacing S is related to the misfit angle ϕ by[2]

$$S = b/\phi \tag{2.20}$$

W_D is found directly from the value of m. Estimates for gallium based on a 1200°C, 20 minute diffusion give $m = 9 \cdot 10^4$, which appears to be too low. A 1050°C phosphorus diffusion for three hours gives $m = 1 \cdot 7 \cdot 10^5$, which is of the same order of magnitude as that found by the "spike velocity" method.[2]

The importance of enhanced diffusion along dislocations with respect to device reliability becomes quite significant in transistors where thin base layers are employed, since a slight nonuniformity in the diffusion fronts of the base emitter and base collector junctions might cause excessive heating due to thin spots,* and eventual failure. Several attempts were made to determine the uniformity with which such diffusion fronts "collide", by observing the current-voltage characteristic between the emitter and collector of a normal transistor structure. This was done immediately after a normal diffusion, and then after several short additional diffusions which should eventually cause a front "collision." Thin regions of the base layer should appear first, as an ohmic connection between the emitter and collector. These measurements were hampered by excessive junction leakage currents, and no conclusions could be drawn concerning this phase of the conductivity measurements.

* See Section II of this report.

The conductivity measurements described in this section have given considerable information which leads to a more thorough understanding of diffusion along dislocations. These measurements, made on structures utilizing a small angle grain boundary, permit analysis of enhanced diffusion from the standpoint of isolated dislocations. The first two configurations utilizing "meeting" diffusion spikes give information, for example, about the enhanced conductivity one might expect as a result of diffusion into dislocations located in the thin base layer of a transistor, causing device failure. The third structure involving the measurement of a "transverse" conductivity vs stepwise etching has provided a new tool for evaluating diffusion data already obtained by other methods. This is particularly valuable in the case of gallium, which gives anomalous results shown by staining. We confirm these anomalous results by our conductivity measurements, indicating that the staining technique is not entirely at fault. The diffusion of acceptors such as gallium seems to be to a large extent uncontrolled. Impurity atmospheres surrounding dislocations is an attractive explanation for the observed grain boundary diffusion anomalies. Electrical conductivity measurements provided an additional method for better understanding the role of dislocations in device failure.

(3) Grain Boundary Devices

Some silicon p-n junction devices were fabricated from grain boundary material. This was done in order to study the effects of the dislocations on the electrical performance of the device, in particular the influence upon the current-voltage characteristic of the junction. *

Shockley[13] has suggested that the dislocations in the group IV semiconductors involve broken unsaturated or "dangling" Si-Si bonds. If this assumption were true, considerable electrical effects should be expected. The most important consequence for practical reliability problems would be the extra conductivity which dislocations with dangling bonds should provide in reverse biased p-n junctions. Earlier measurements (as quoted in Ref. 2) had already indicated that these effects were much smaller than expected.

Some further investigations were conducted under the present contract. Diodes with small angle grain boundaries were constructed for this purpose, (see Fig. 6, Appendix). Bicrystal slices were diffused with phosphorus, boron and gallium. Mesas were masked with a wax spraying technique and then etched. The masks were aligned such that a maximum number of diodes contained the boundary. The boundary dislocations ran perpendicular to the diffusion front. For several experiments, contacts were applied by nickel plating, and the diced diodes were put in a glass package. This made it possible

---

* For a further discussion of these problems, also see the Appendix.

[13] W. Shockley, Phys. Rev. 91, 228 (1953).

to age the diodes at elevated temperatures. Other mesa diodes were left on the supporting slice undiced; there the electrical measurements were done by means of four pressure-contacts. Forward and reverse current-voltage characteristics were checked. The results obtained are discussed below.

a. Reverse Characteristics

The most important results of the various series of measurements can be summarized as follows: Diodes which were diffused under clean conditions, especially under "gettering" conditions[14] with glassy oxide layers showed no influence of the dislocations. Grain boundary diodes were "hard", i.e. the reverse current densities remained low up to the breakdown voltage. No difference was found between these devices and those away from the boundary. On the other hand, diodes which were deliberately contaminated (in particular with metals, such as done by Goetzberger and Shockley[14]) showed "softness" with the grain boundary diodes being particularly susceptible to this "softness".

These results indicate that the direct influence of dislocations is small. The high reverse currents expected from dangling bonds are not observed. Either the dangling bonds are avoided by a diffusional rearrangement (see Fig. 3, Appendix) of the silicon atoms at the dislocation core or the dangling bonds are saturated by some impurities. However, the dislocations seem to provide favorable nucleation sites for the formation of metal precipitates, which are known to cause softness in diodes

[14] A. Goetzberger and W. Shockley, 31, 1821 (1960).

Some of the grain boundary diodes were aged at elevated temperatures together with control diodes from the same slice without the small angle grain boundary. The results of one run are shown in Fig. 2.7. These n^+p diodes were fabricated by a phosphorus diffusion into 0.4 ohm cm boron-doped bicrystal material. Aluminum contacts were evaporated and alloyed. It is seen from the curves of Fig. 2.7 that the aging did not lead to catastrophic failure. The final aging step (curve e) even produced a more favorable characteristic than the initial one. Apparently surface effects are dominating and the influence of the grain boundary dislocations seems negligible. However, in some runs the boundary diodes appeared to have superior characteristics after aging as compared to the diodes off the boundary. One might hypothesize an "internal gettering effect", but statistical evidence is yet too small to warrant any such statements.

b. Forward Characteristics

Grain boundary diodes which were produced by boron or phosphorus diffusion did not reveal any irregularities in their forward current-voltage characteristics. A few measurements on Al-diffused diodes, however, showed that there is an effect exerted by the grain boundary.

The forward characteristic of a silicon diode is usually written in the form:

$$I = I_o \exp\ [\ (qV/AkT) - 1\ ], \qquad (2.21)$$

where I and V are the current and voltage, T the absolute temperature, k the

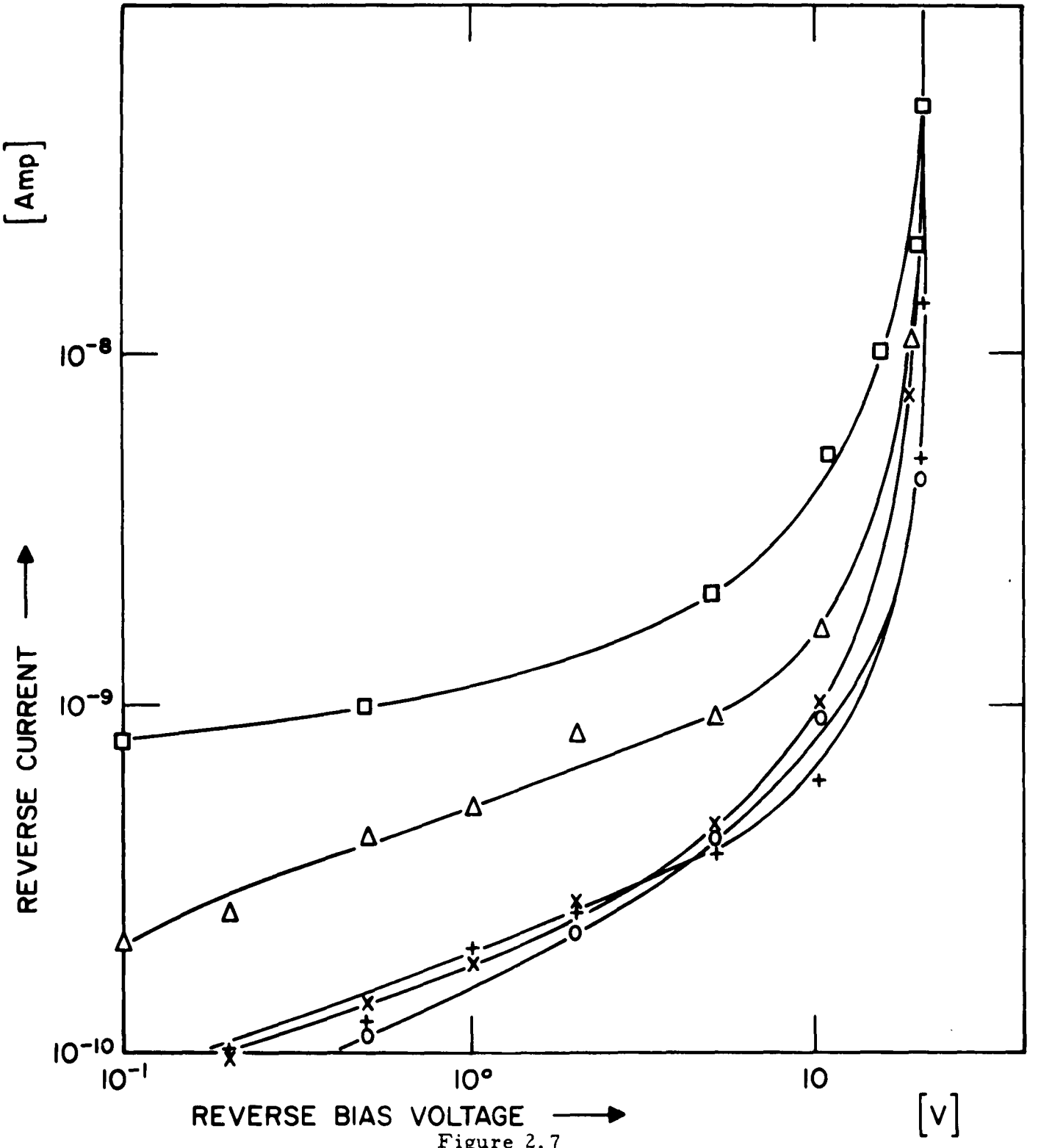


Figure 2.7

Aging behavior of n^+p-mesa diodes with a small angle grain boundary perpendicular to the diffusion front

a) △ Before aging
b) + 16 hours at 175°C
c) O 64 hours at 175°C
d) □ 4 hours additional at 260°C
e) x 20 hours additional at 260°C

10-18-62
50408

Boltzmann constant; I_o and A are parameters. A should normally fall between 1 and 2[15]. Certain anomalies have frequently been observed, especially in silicon solar cells[16]. These devices require the assumptions of an A-value of greater than 2, which is difficult to explain theoretically. Surface effects could be the underlying physical phenomenon[17], but there is also some evidence that impurities, in dissolved or precipitated form could play a role[18]. The hypothesis was offered[18] that metallic impurities in precipitated form within the space charge layer cause the anomalously high values of A through some as yet unknown shunting mechanism.

The aluminum-diffused diodes at the grain boundary consistently showed values of A≈3 and even higher. The Al-diffused diodes away from the boundary had A-values slightly below 2, as theory predicts. It is tempting to interpret this result as an affirmation of the precipitate hypothesis. Aluminum is known to precipitate easily at dislocations.[12] The diffusion spikes obtained after the aluminum diffusion had a spotty apearance, which would indicate loss of Al-acceptors through precipitation into a metallic phase. The grain

---

[15] C. T. Sah, R. N. Noyce and W. Shockley, Proc. IRE 45, 1228 (1957).

[16] W. Shockley and H. J. Queisser, J. Appl. Phys., 32, 510 (1961).

[17] C. T. Sah, Proc. IRE ED-9, 94 (1962).

[18] H. J. Queisser, Final Report Contract AF33 (616)-7786, also Solid State Electronics 5, 1 (1962).

boundary dislocations provide an excellent means of inducing precipitation in a reasonably controlled fashion. Further such studies should provide more information to solve the problem of the mysteriously high A-values. This question is of considerable importance for the basic understanding of p-n junctions (particularly for silicon solar cells); the reliability aspects, however, are less pertinent. Therefore it was decided not to investigate this problem in more detail under the present contract.

In the Quarterly Report #1 for this contract it was reported that Al-diffused diodes at grain boundaries exhibited a "seemingly negative resistance" in the forward direction. Four point measurements had shown a region of decreasing potential with increasing current. This is, however, only observed if both pairs of contacts are placed on opposite grains of the bicrystal diode. Under these conditions one has two diodes in parallel, connected by two resistors (across the boundary). The resistance across the boundary seems especially high after the Al-diffusion treatment. Oxide precipitation is a likely cause of this. Thus, the early hypothesis of an injection phenomenon from Al-precipitates appears to be incorrect; the effect could simply be explained by the unique circuit conditions involving diode and resistor pairs.

B. Photoresponse Measurements

When a light spot is scanned across a grain boundary of a semiconductor, there is a behavior similar to that of a p-n junction photocell; electron-hole pairs are created, the base material collects one type of carrier and the grain boundary the other.[19] This results in a photovoltage, and the apparent conductivity type of the grain boundary is identified by that carrier which it collects. This photoscanning technique has been used quite extensively with germanium bicrystals.[19, 20] We report photoscanning measurements on silicon bicrystals containing a small angle grain boundary, and suggest that impurity atmospheres account for the electrical behavior of these grain boundaries. Since drastic changes in the photoresponse appear after heat treatments which are known to lead to phase changes of oxygen in silicon, we suspect an influence of oxygen atmospheres on the electrical properties of the dislocations.

When a germanium bicrystal is photoscanned, the grain boundary collects holes, and therefore appears to be a p-type layer. This behavior is true regardless of the conductivity type of the base material, or impurity atmospheres at the grain boundary. The explanation given is the acceptor nature of the dangling bonds at the grain boundary dislocations.[20]

---

[19] G. L. Pearson, Phys. Rev. 76, 459 (1949).

[20] O. Weinreich, et al, Brussels Conf., 1958 in Solid State Physics (Academic Press, Inc., New York, 1960) Vol. 1, Part 1, 97.

Silicon, however, behaves quite differently. A few photoscanning measurements by Weinreich[20] and Maturkura[21] indicate that a grain boundary in silicon collects electrons, and therefore appears to be an n-layer. We find similar results, and report more extensive photoscanning measurements on silicon bicrystals containing a 10^{o} angle grain boundary. A low angle of misfit was chosen to insure that the boundary was made up of distinct and isolated edge dislocations. In our samples the dislocation spacing based on a 10^{o} angle of misfit is approximately 4 lattice constants, or about 20^{o} A. A diagram of the apparatus used for these photoscanning measurements is shown in Fig. 2.8. The equipment was designed and built by R. Haitz of our laboratory for microplasma studies under Cambridge Research Center Contract #AF19 (604)-8060. The apparatus was modified for our measurements.

The light from a high pressure mercury arc lamp (PEK 109) is focused on a "chopper" disc mounted on a synchronous motor. Holes are cut in the disc such that during operation approximately 390 light pulses per second are produced. The light pulses are focused through a lens system onto a pin-hole aperture mounted on a microscope. The image of the illuminated aperture is formed by the microscope objective lens as a small light spot on the microscope stage, reduced in size from the aperature diameter by an amount equal to the objective lens power. By this arrangement light spots as small as 1μ in diameter have been produced. The half silvered mirror

---

[21] Y. Matukura, J. Phys. Soc. Japan 16, 842, (1961).

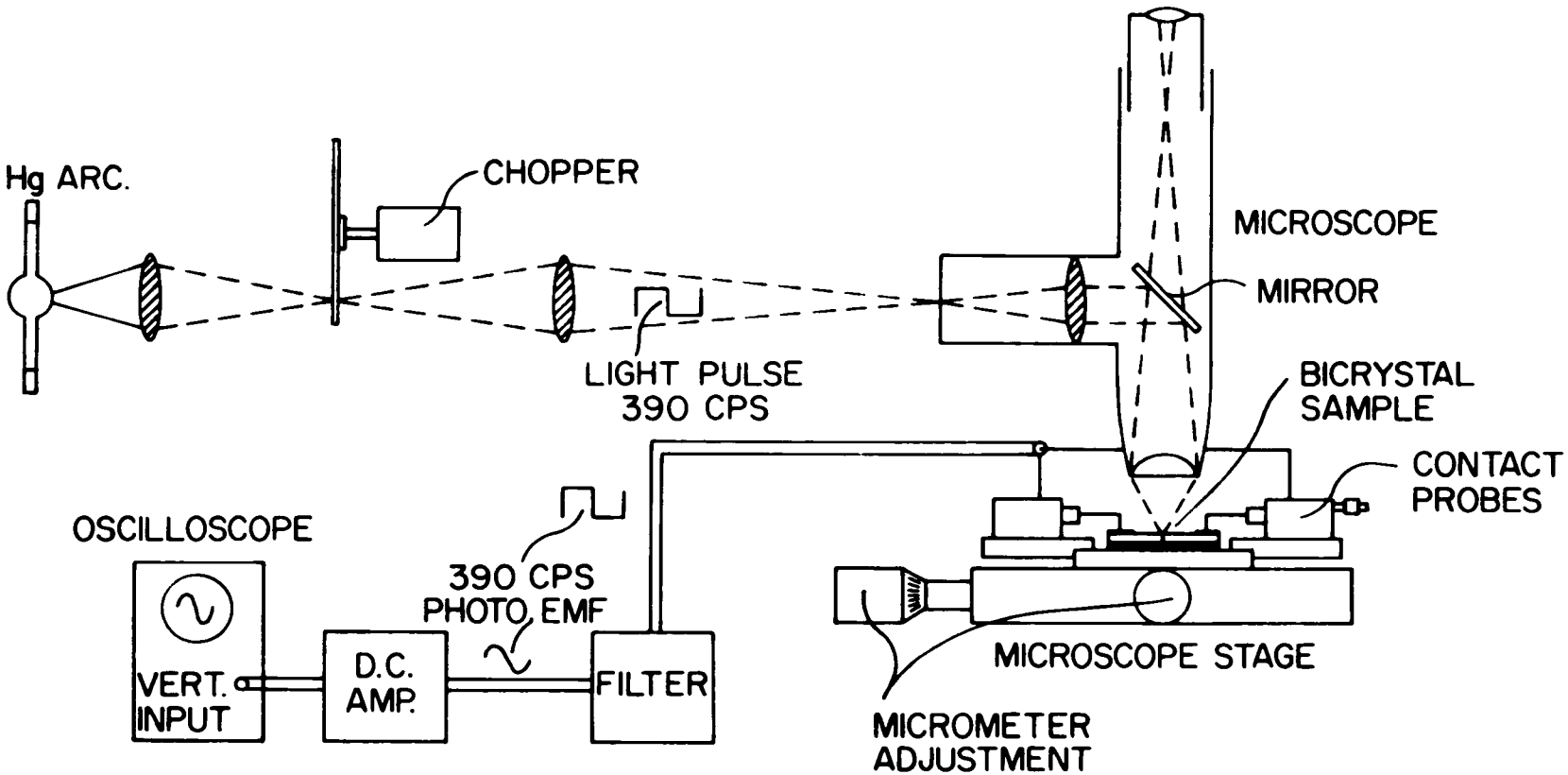


Figure 2.8

Apparatus used for photoscanning measurements

in the microscope allows viewing of the light spot formed on the microscope stage.

The grain boundary sample to be photoscanned is placed on the microscope stage. Electrical contact to the sample is made by small probes on evaporated gold strips at each side of the grain boundary. The probes and stage are an integral unit, so that motion of the stage does not involve lifting the probes from the sample. Motion in a horizontal plane is calibrated by micrometer screw adjustments.

When the grain boundary is illuminated by the light spot, a photovoltage is produced at the contacts, in the form of a square pulse of 390 cps repetition rate. The pulse is sent through a narrow band filter (±1 cps bandwidth) which is tuned to 390 cps, producing a 390 cps sinusoidal signal. This filter eliminates signal interference from 60 cps and its harmonics. The filtered 390 cps photovoltage is then amplified 1000X by a Tektronix type 122 d.c. amplifier, and displayed on an oscilloscope. This relatively elaborate filtering and amplification arrangement was found necessary since the grain boundary photovoltages were in the 1-10 μV range.

The samples used for these photoscanning measurements were slices approximately 200μ thick, cut from bicrystals. Oxygen-rich bicrystals were grown from a quartz crucible; "oxygen-free" bicrystals were obtained with the pedestal method to be described later. The grain boundary edge dislocations in all cases run perpendicular to the surface of the slice.

Several crucible grown bicrystals, n- and p-type, varying in resistivity from 0.4 to 2000Ω cm have been photoscanned with the apparatus previously described. Photoresponse has been found only with high resistivity (1000-2000Ω cm) p-type (boron-doped) as grown bicrystals. High resistivity (600-1000Ω cm) "oxygen-free" as-grown bicrystals did not show any photoresponse. Fig. 2.9 shows a plot of the photovoltage vs distance of the light spot from the grain boundary for a 2000Ω cm p-type crucible as-grown bicrystal. The plotted voltage is D.C., and the magnitude obtained from the peak-to-peak A-C. reading of the 390 cps signal displayed on the oscilloscope. Scanning was done with a light spot 25μ in diameter in most cases. As the light spot is scanned from the grounded side of the sample the photovoltage first increases negatively, reaching a maximum value when the spot is tangent to the boundary. As the spot is symmetrically over the boundary (x = 0) the photovoltage drops to zero and then increases to a maximum positive value when the spot is tangent to the boundary on the opposite side.* This polarity relationship indicates that the silicon grain boundary collects holes and therefore acts as an n-layer.

---

* The dependence of the photovoltage upon distance X from the collecting grain boundary is exponential, as shown in Fig. 2.10. The decay is of the form exp (-X/L), where L is the diffusion length of the carriers. It can be expressed as $L = \sqrt{(\mu kT\tau)/q}$, where μ is the minority carrier mobility, τ their lifetime, q their charge, k is Boltzmann's constant, T is the absolute temperature. Since L is measured with this technique, either μ or τ can be obtained when the other one of these coefficients is known. This was not the case in these experiments, therefore no such determination was made. However, the values of L obtained are not in disagreement with reasonable assumptions for mobilities and lifetimes.

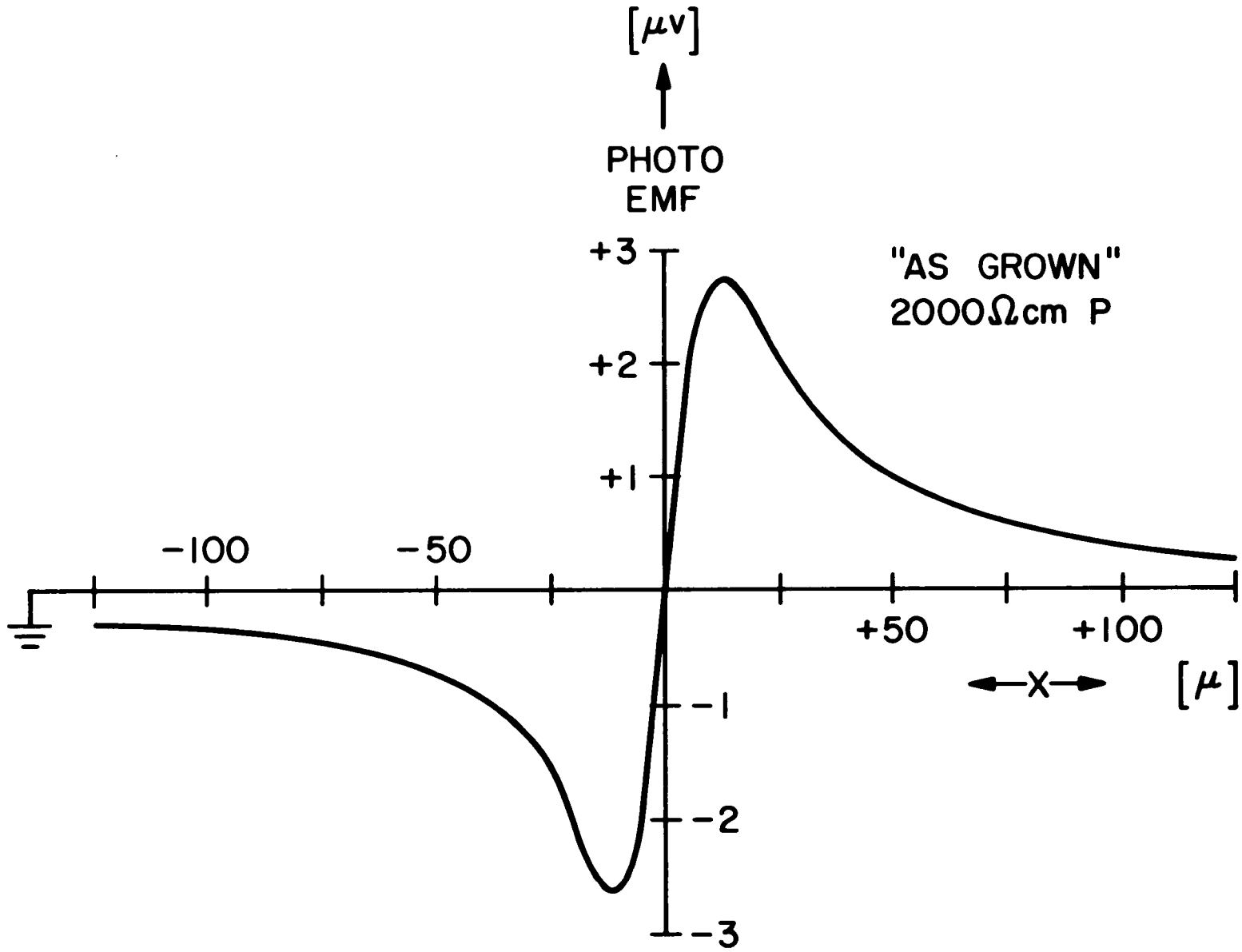


Figure 2.9

Photovoltage vs light spot position of 2000 ohm cm p-type "as grown" bicyrstal

2-26-62
50147

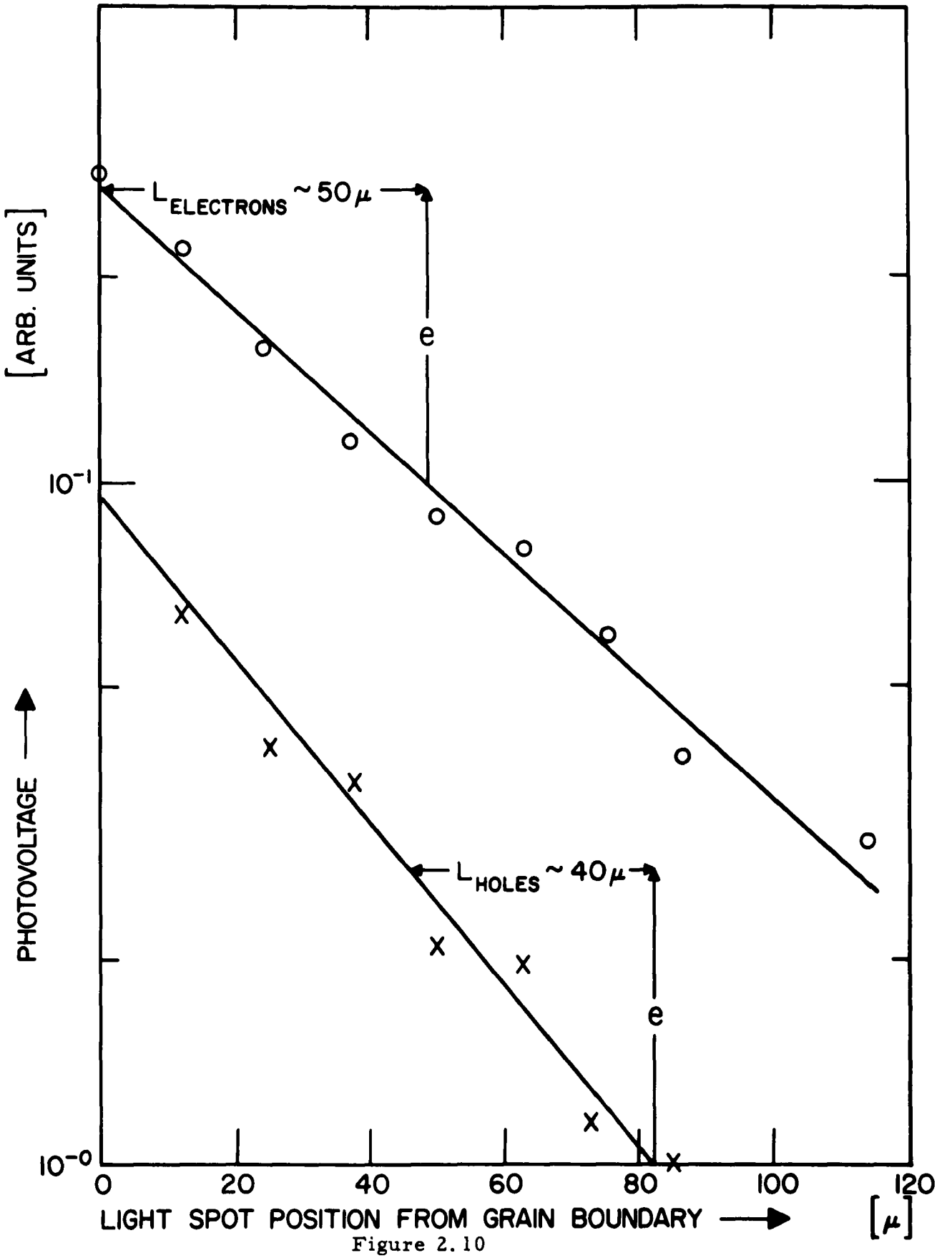


Figure 2.10

Dependence of photovoltage with light spot distance from grain boundary

10-17-62
50405

If impurity atmospheres are responsible for the photovoltage, one should see an influence of the thermal history of the sample. The samples were therefore subjected to various heat treatments. These heat treatments were carried out after the gold contacts were removed in aqua regia to prevent gold contamination. In one particular experiment, the sample was broken into two parts, perpendicular to the grain boundary, permitting different heat treatments on the same sample. Both halves were heated in a hydrogen atmosphere at 1100°C for 15 minutes. One part was cooled by quenching to room temperature in silicone oil, the other part cooled at 150°C/minute. The gold contacts were reapplied, and photoscanning repeated. Fig. 2.11 shows a plot of photovoltage vs light spot position from the boundary for this heat treatment. It is observed that the polarity of the photovoltage is reversed. This polarity change indicates that the grain boundary now collects holes and appears as a p-layer, opposite to the n-type behavior of the as-grown material. In addition to the p-type conversion of the grain boundary, the magnitude of the photovoltage is dependent upon the cooling rate. That part which was cooled at 150°C/minute shows twice the photovoltage of that which was quenched. An additional sample with the same as-grown photoresponse characteristics was heated for four hours at 405°C in

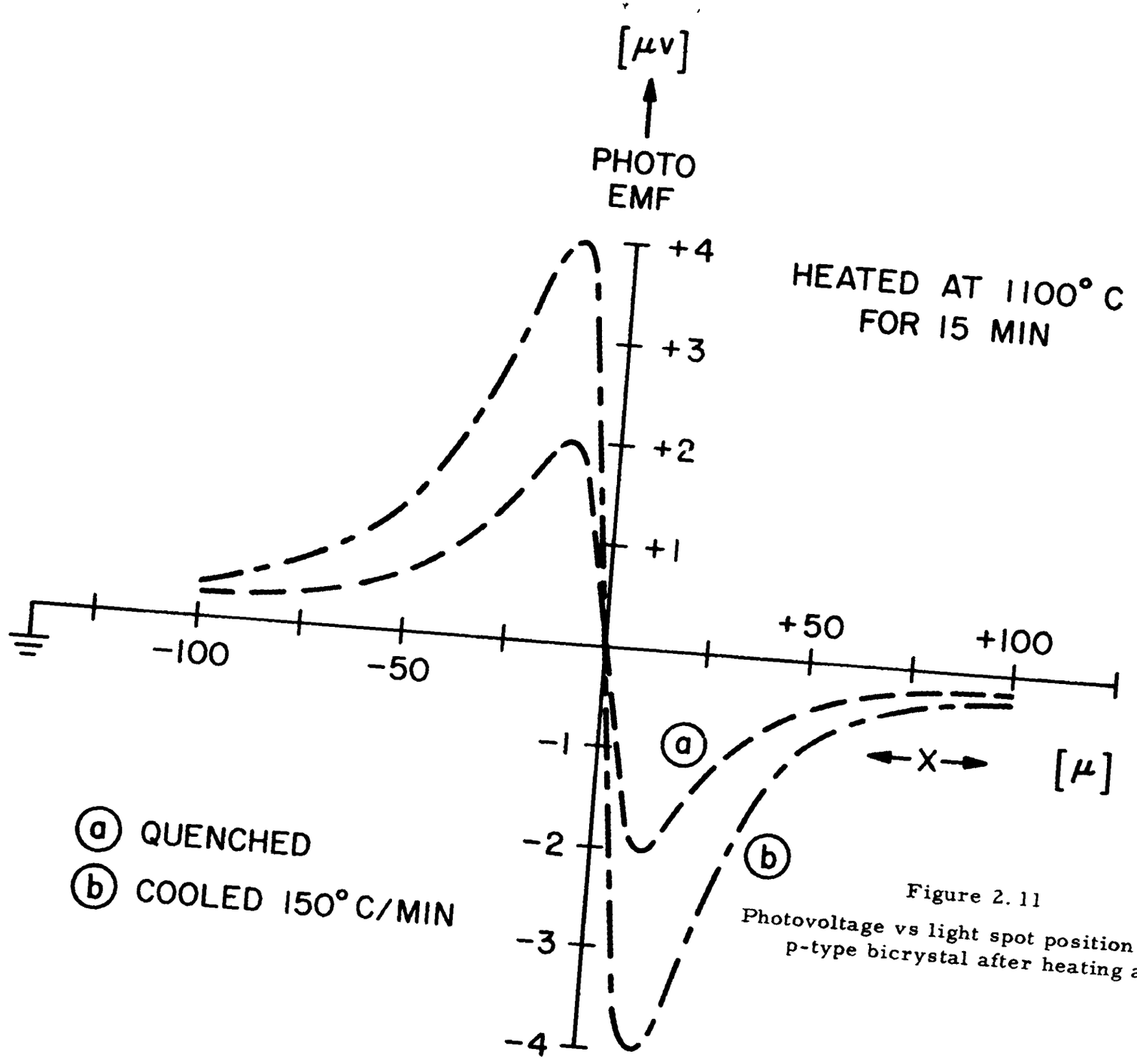


Figure 2.11

Photovoltage vs light spot position of 2000Ω cm p-type bicrystal after heating at 1100°C

2-26-62
50148

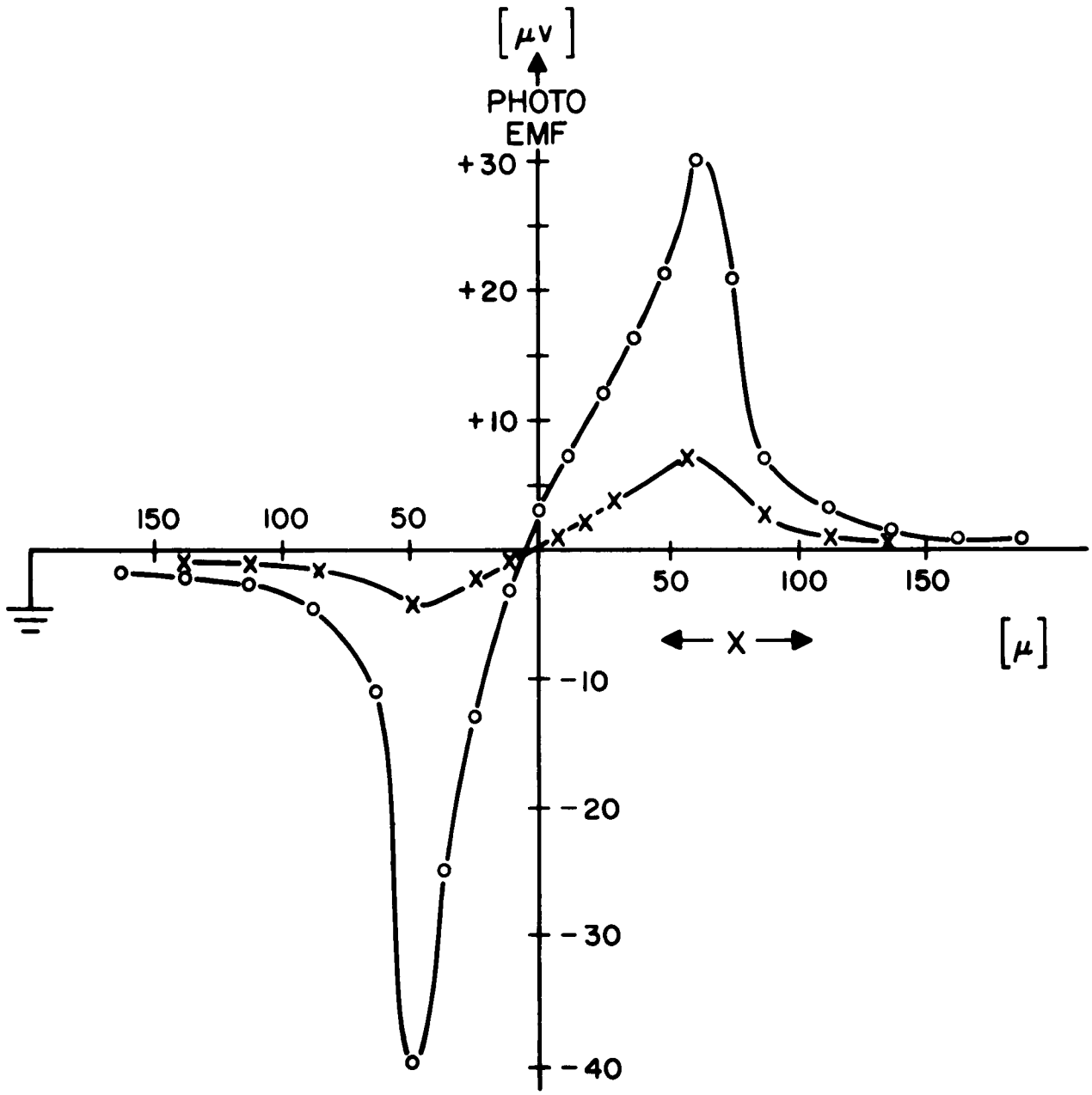


Figure 2.12

Photovoltage vs light spot position of 2000Ω cm p-type bicrystal before and after heating at 405°C

hydrogen atmosphere, and quenched. The photoresponse of this sample before and after heating is shown in Fig. 2.12. Although there is a reduction in the magnitude of the photovoltage, there is no conversion to p-type behavior as there is following the 1100°C treatments. As mentioned earlier, high resistivity p-type "oxygen-free" as-grown bicrystals show no photoresponse. Following a 15 minute heat treatment at 1100°C in hydrogen atmosphere, however, they yield a p-type photoresponse.

Bicrystal samples which showed a photoresponse conversion from n- to p-type were examined by staining before and after heating. This staining method is commonly used for p-n junction delineation.[9] Fig. 2.13 shows staining pictures of one sample before and after heating at 1100°C. A portion of the sample containing the grain boundary was beveled at a small angle, and a staining solution applied under strong illumination. The staining solution consisted of 10 cc H_2O, 10 cc HF acid, and 5 drops HNO_3 acid. Those areas which take on a dark stain are p-type, those which remain light are n-type. In the "as grown" sample there can be seen an n-type "core" along the grain boundary surrounded by two p-type "tracks". After heating at 1100°C, the grain boundary region has taken on a p-type stain. This staining result seems to be in agreement with our photoscanning measurements.

Samples which yielded an n- to p-type conversion after heating at 1100°C were reheated at 1300°C for one hour and quenched,

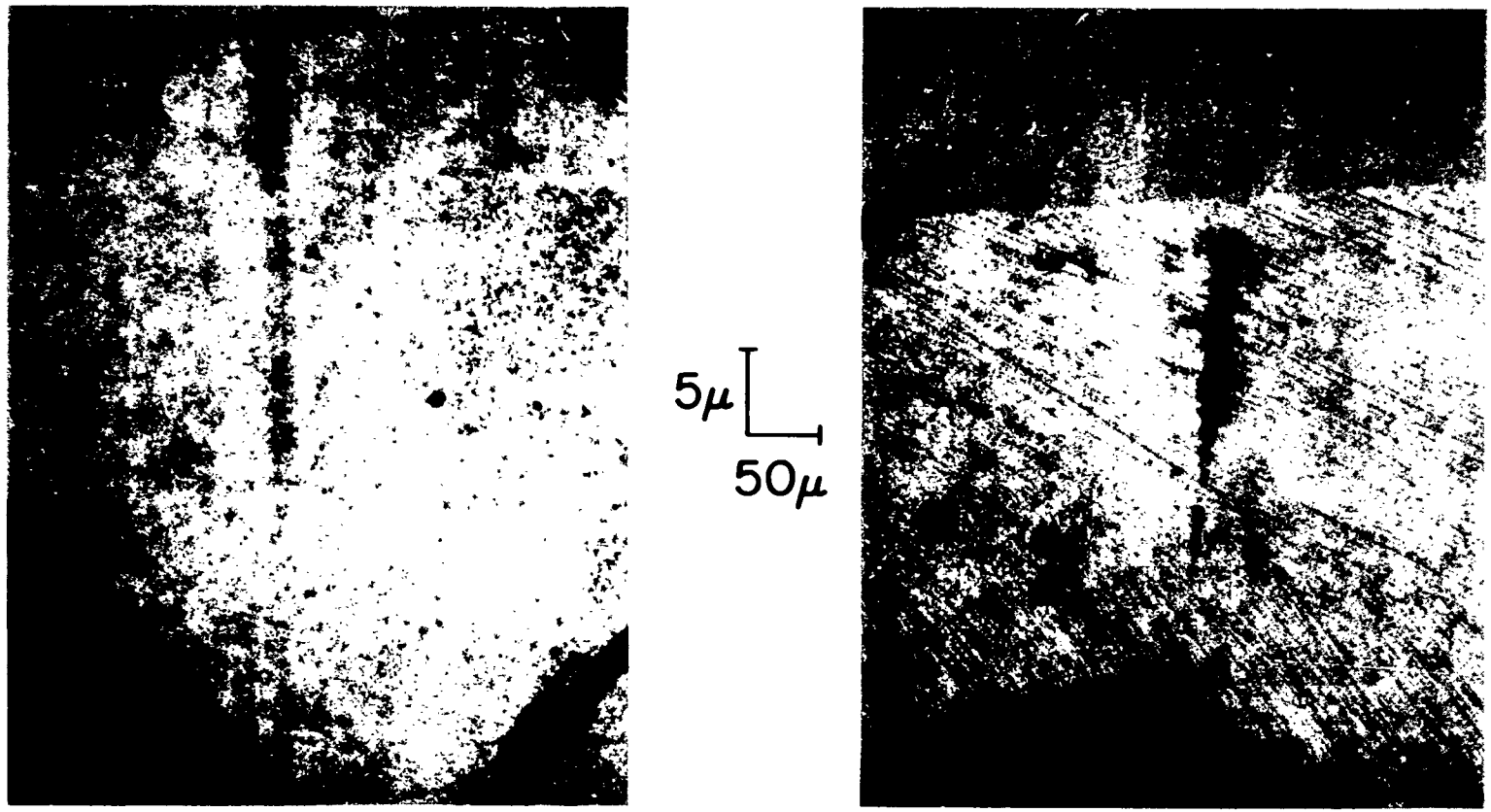


Figure 2.13

Photographs of stained bicrystal showing n-type boundary converted to p-type after heating

The 1300°C treatment almost completely annihilated the photoresponse. A subsequent heating at 450°C for 4 hours had no further effect.

In addition to experiments involving heat treatment in a "clean" atmosphere, several bicrystal samples were diffused with both p- and n-type impurities (Ga, B, P) using standard techniques for p-n junction fabrication. Following these diffusions, the samples were etched to remove the diffused surface layer. Photoscanning these diffused samples reveals that the grain boundary always acts like a p-layer after donor and acceptor diffusions. These results seem surprising since the nature of the diffusant seems to be of no consequence. Apparently the decisive factor in these experiments is the heating associated with the diffusions; in all cases the samples were heated to temperatures between 1000°C and 1200°C.

This photoelectric behavior of silicon grain boundaries is quite unlike that reported for germanium. We suggest an explanation of the phenomena with the assumption of impurity atmospheres rather than hypothesizing dangling bonds. Lack of observation of both spin resonance[22] and conduction effects[23] is evidence against the dangling bond model in silicon in its simple form.

On the other hand, the photoresponse data indicate the action of impurities, since there is an obvious dependence on temperature and

---

[22] W. T. Read and G. L. Pearson, Report Bristol Conf. Defects in Crystal Solids, 1954 (Phys. Soc., London, 1955) p. 144

[23] See section on "Grain Boundary Devices" and Appendix, this report.

rates of temperature change. The hypothesis of oxygen impurity atmospheres seems particularly attractive. Oxygen is known to be present in crucible-grown silicon in quantities as high as 10^{18} cm^{-3} It can exist in several different configurations in the lattice[24]: dissolved, precipitated as SiO_2, or in a donor complex SiO_n^+.

The transformations between these phases have been studied extensively.[24] It is known that there is precipitation of SiO_2 at about 1000°C. This reaction could have a relation to our observed polarity changes towards p-type boundaries. Heating at 1300°C redissolves the precipitates. This could be the explanation for the annihilation of the photoresponse after the heat treatment at 1300°C and subsequent quenching. Donor complex formation predominates around 450°C; our experiments do not reveal a clear correlation with this process, however. The absence of an initial photovoltage in the supposedly cleaner and oxygen-free samples also indicates an influence of impurities upon the photoresponse.

The experimental facts seem to favor an explanation based upon impurity atmospheres. Thus far there is, however, only indirect evidence for the influence of oxygen impurity atmospheres. Other authors made similar assumptions. Lederhandler and Patel[25] reported that there is preferential precipitation of oxygen at dislocations.

---

[24] For a review, see W. Kaiser, Semiconductor Conference, Prague, 1960.

[25] S. Lederhandler and J. R. Patel, Phys. Rev. 108, 239 (1957).

Pearson, et al[26] in their studies on fracture and deformation of silicon have also postulated that dislocations are affected by impurity atmospheres and constitute preferential sites for SiO_2 precipitation. The present photoresponse data are not in disagreement with these concepts. It is, however, possible that other impurities such as metals or carbon determine the electrical effects of the dislocations, perhaps in combination with the oxygen present in the crystals. Photoresponse measurements could provide a useful tool for future investigations of these problems.

---

[26] G. L. Pearson, W. T. Read, and W. L. Feldman, Acta Met. 5, 181 (1957).

C. Growth of Silicon Bicrystals by the Dash Pedestal-Method

(1) Introduction

Small angle grain boundaries consist of a regular array of edge dislocations[27, 28, 29]. Bicrystals with such boundaries are an excellent tool to investigate dislocations. From the result of grain boundary experiments one can draw conclusions about the behavior of isolated dislocations.

Silicon bicrystals have been studied in the past[2, 21, 30, 31, 32]. Enhanced diffusion along small angle grain boundaries in silicon has been found in this laboratory[2]. Phonon drag measurements in silicon grain boundary specimens have been carried out by Hubner and Shockley,[30] and light emission at high voltage from a grain boundary junction has been observed by Goetzberger and Stephens[31]. All these experiments were made with oxygen-rich bicrystals grown from a quartz crucible by the Czrochralski method[7]. It is, however, not well known what role oxygen plays in the behavior of silicon crystals together with dislocations.

---

27 W. L. Bragg, Proc. Phys. Soc., (London) 52, 54 (1940).

28 J. M. Burgers, Proc. Phys. Soc., (London) 52, 23 (1940).

29 Summary of experimental results, S. Amelinckx and W. Dekeyser, in Solid State Physics, ed. F. Seitz and D. Turnbull, (Academic Press, New York and London, 1959), Vol. 8, p. 420.

30 K. Hubner, W. Shockley, Advanced Energy Conversion 1, p. 93 (Pergamon Press, Great Britain, 1962).

31 A. Goetzberger, C. Stephens, J. Appl. Phys. 32, 2646 (1961).

Hooper and Queisser[32] recently found a strong influence of heat treatment on the photoresponse of Si bicrystals. They assumed that oxygen particularly accounts for these effects. For further studies it is, therefore, desirable to have both oxygen-rich and oxygen-free silicon grain boundary crystals available. All bicrystals obtained by the standard crucible-procedure are contaminated with oxygen. This article describes a method and an apparatus for growing oxygen-free silicon bicrystals.

(2) Method of Growing Dash Bicrystals

Dash[33, 34, 35] describes a method of growing silicon mono-crystals free of dislocations and oxygen. This procedure does not employ the quartz (SiO_2) crucible necessary with the well-known Czochralski method.[7] The use of quartz crucibles gives rise to high oxygen content in the pulled crystals. With the Dash method, the crystal is pulled from the molten top of a silicon rod, commonly called the "pedestal". A seed consisting of material of low dislocation density is used. It is tapered at the tip to a diameter of approximately 0.1 mm. The crystal is pulled in an inert argon atmosphere to avoid contamination and oxidation.

---

32 W. Hooper, H. J. Queisser, Bull. Am. Phys. Soc. 7, 211 (1962)

33 W. C. Dash, J. Appl. Phys. 29, 736 (1958).

34 W. C. Dash, J. Appl. Phys. 30, 459 (1959).

35 W. C. Dash, J. Appl. Phys. 31, 736(1960).

For these experiments, a Dash crystal-growing apparatus was constructed. After many perfect silicon monocrystals had been grown, the equipment was used to grow grain boundary crystals with low oxygen content. Silicon material containing grain boundaries was chosen for the seeds, instead of the dislocation-free bars employed in the conventional Dash method.

a. Crystal Grower

A Lindberg float zone scanner (Model LA-VSE-24 ZR) was converted into a crystal grower. The basic arrangement is shown in Fig. 2.14.

The quartz tube, L, contains the main components of the crystal grower. The tube (35 mm O.D.) is aligned inside the heating coil, N, and shading coils, M, of the Lindberg RF generator (Model LI-10-C-1, 4MC 10 KW). During growth, argon gas containing less than 0.1 ppm of oxygen flows through the quartz tube.

The pedestal, F, is connected to the quartz chuck, H, (15 mm diameter, 100 mm length). A slotted steel tube, J, with centering screws supports the quartz rod. The top of the rod is cut to a step shape to fit a corresponding step on the pedestal material. The parts are connected by tantalum wires. One of these wires, G , establishes the coupling of the RF field before the silicon pedestal is heated. The quartz chuck, H, insulates the hot parts of the grower. It also helps to reduce contamination of the pedestal, and improves the ease of aligning

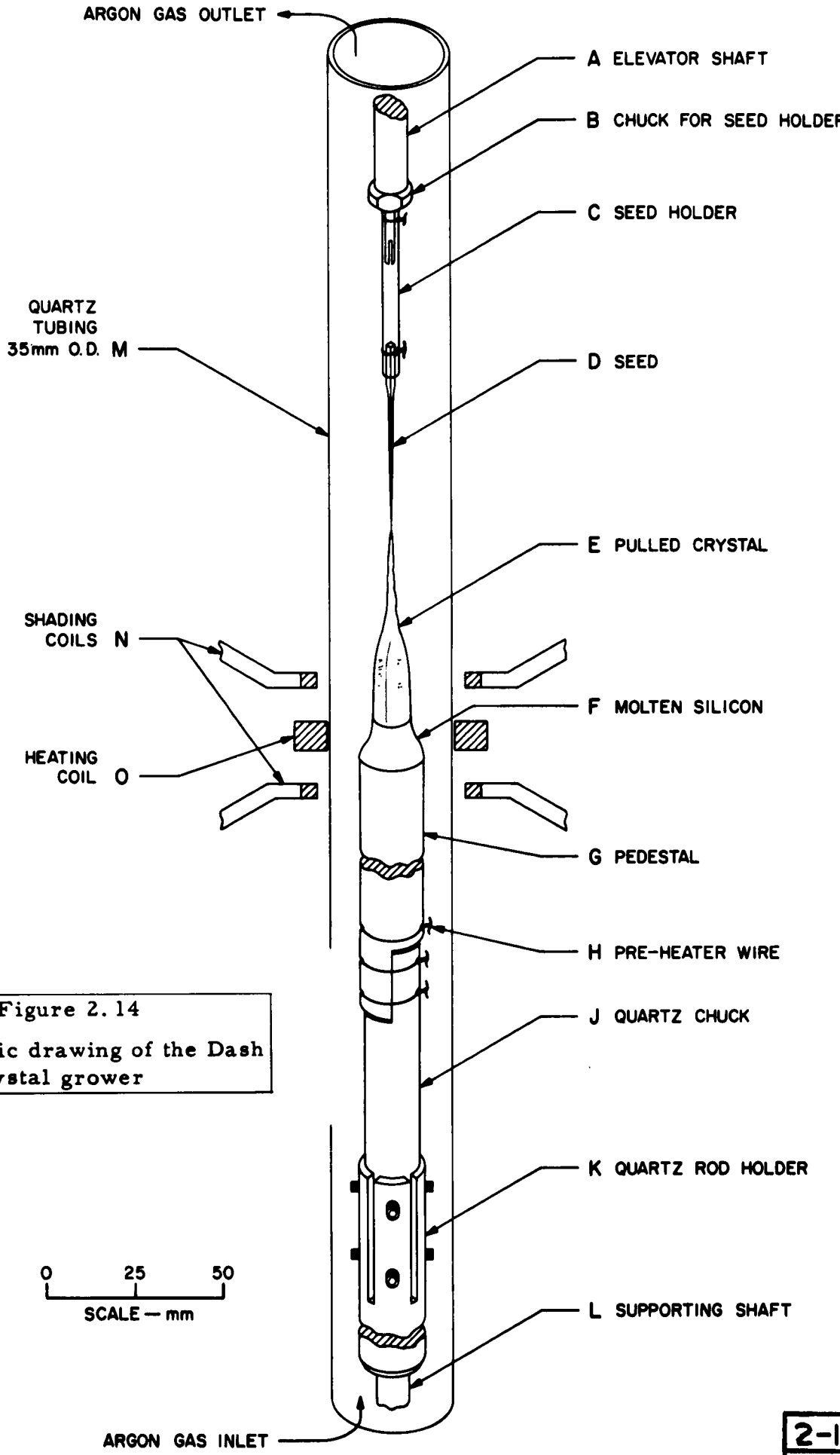


Figure 2.14

Schematic drawing of the Dash crystal grower

2-19-62
50141

the whole assembly. The supporting shaft, K , is mechanically connected to a motor unit. Thus the operator is able to rotate and move all parts, F-K, according to any desired program.

A small quartz tube, B , is used as a "seed holder". It is connected to the elevator shaft A. The crystal, D , is pulled from the melt, E , by moving the seed, C , upwards. At the same time it rotates with a constant speed of 6 rpm. The solid liquid interface is kept at a constant level by feeding the pedestal upwards to compensate for the removed material.

b. Preparation of Silicon Starting Material

Czochralski-grown crystals (25-500 ohm cm, N-type and 7-15 ohm cm, P-type) with a diameter of about 18 mm were used as pedestal materials. The top portions of all these rods were tapered to a cone with a SiC-abrasive, while the lower parts were cut to fit the quartz chuck. Subsequently, the pedestal was etched in a solution consisting of 1 part HF (49%) and 3 parts of HNO_3 (70%).

The seed was prepared in the following way: A bicrystal containing a 10.5 grain boundary was grown by the Czochralski method using two seeds disoriented with respect to each other (as described by H. F. Matare and H. A. R. Wegener[3]). From this crystal a rectangular piece, containing the grain boundary[36] was cut (2 x 2 x 60 mm). It was

---

[36] H. J. Queisser (unpublished).

lapped by hand until the grain boundary was equidistant from both sides parallel to the grain boundary plane. This "bi-seed" was etched to a needle-like shape (see Fig. 2.15) with the grain boundary still remaining in the center of the tip. The bi-seed must have a small diameter (approximately 0.1 mm) at the tip in order to avoid freezing of the molten silicon on top of the pedestal when the seed is lowered into the melt.

An apparatus for facilitating the etching of the seed crystals was constructed (see Fig. 2.16). The etching apparatus moves the seed vertically, while simultaneously rotating it at 10 rpm. This way, seed crystals can be etched very accurately to the desired concentric geometry. The etching apparatus is similar to a "drill press". The supporting block, A, can be turned about the column, B, so that the seed, E, may be alternatively dipped into the etching solution, C, 1 part HF (49%), 3 parts HNO_3 (70%) or cooling water, D. The seed is connected to the synchronous motor, G, by means of the teflon chuck, F.

(3) Experimental Results

Figure 2.17 is a photograph of the Dash grain boundary crystal. The 10.5° grain boundary is seen on the crystal surface as a fine, sharp line which is not completely parallel to the growth lines.

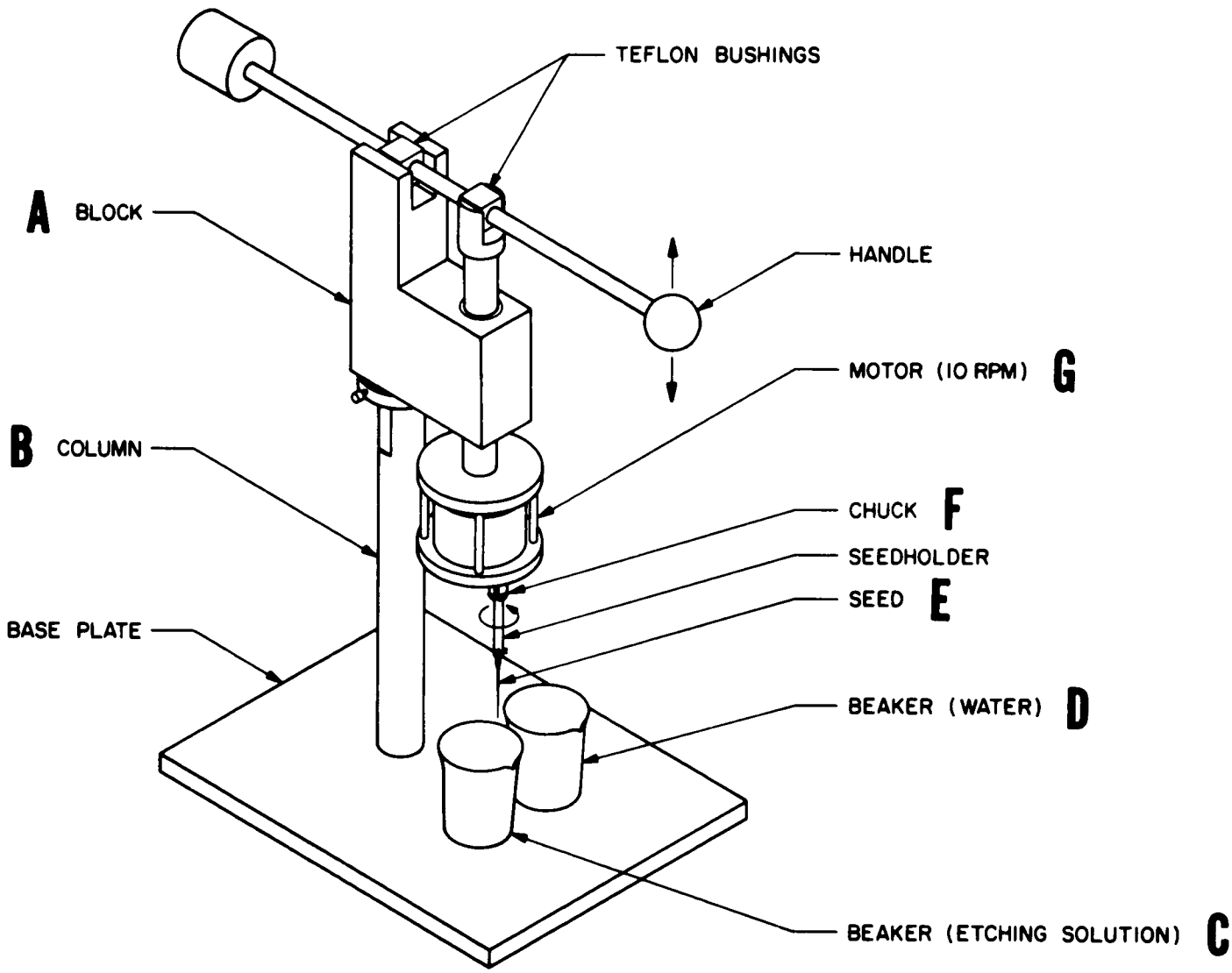


Figure 2.15

Etching Apparatus

4-2-62
50179

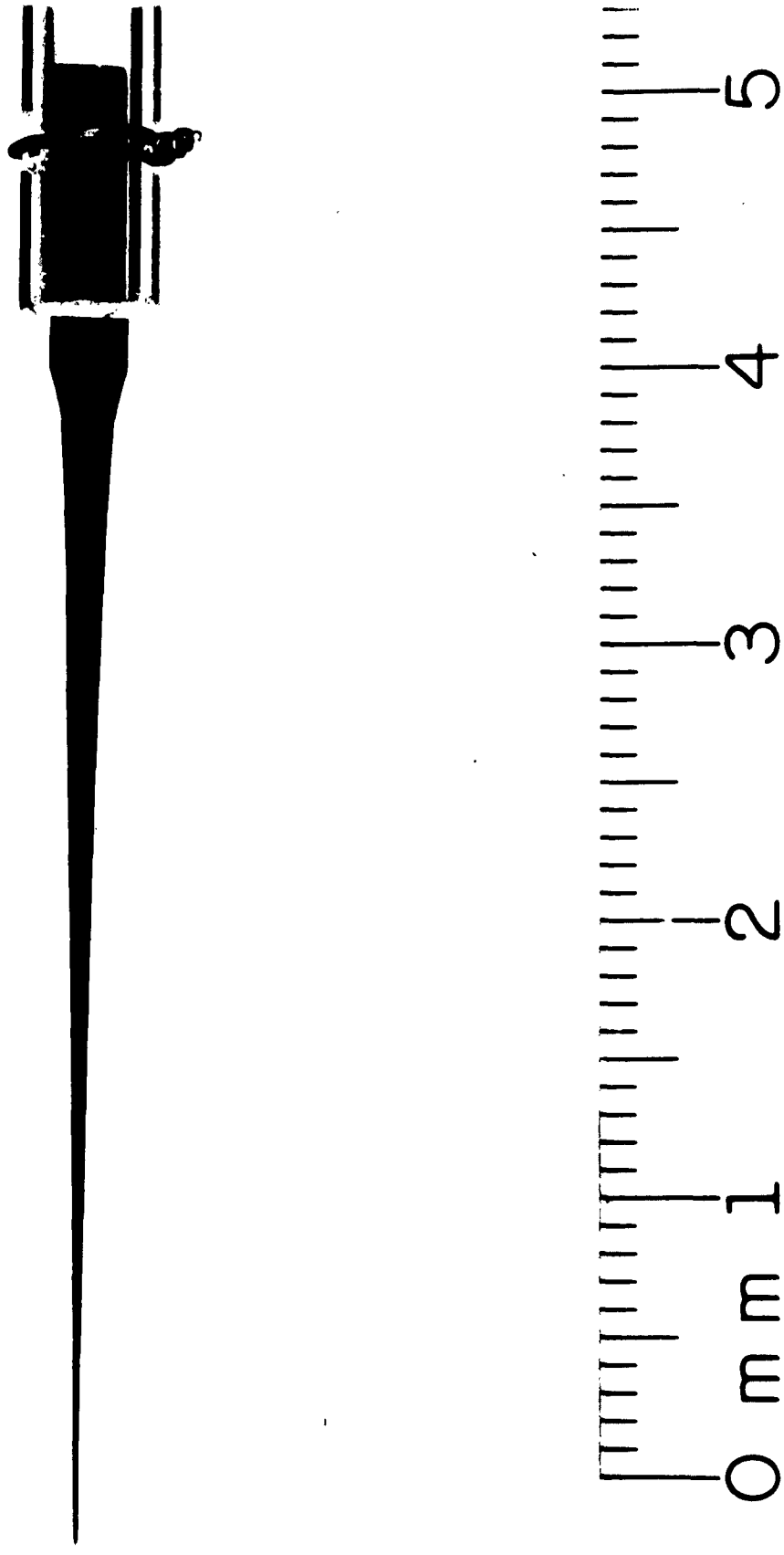


Figure 2.16

Etched bi-seed containing a 10.5° grain boundary (diameter of the tip is about 0.1 mm)

This is caused by inaccurate orientation of the grain boundary with respect to the bi-seed. The crystal is 147 mm long and has a nearly constant diameter of 9 mm over a length of approximately 80 mm. It was grown with a starting pull rate of 4 mm/min. and a final pull rate of 1.4 mm/min. The output energy of the RF generator was manually controlled to achieve optimum growing conditions and the temperature of the molten Si was determined using an optical pyrometer.

Earlier infrared transmission measurements had indicated that Dash crystals grown from unrefined normal Czochralski crystals or from float zone refined silicon bars showed the same absorption peak height at 9μ. The absorption coefficient of all these slices was calculated to be in the order of 0.8 cm^{-1}. This value is known to be the background absorption of silicon itself[37]. All measurements were carried out with a Beckman spectrophotometer (Model IR4). Since this instrument has a double-beam arrangement, the test slices cut from crystals grown by the Dash-method could be directly compared with equivalent float zone material placed in the reference path. A slice taken from the bicrystal shown in Fig. 2.17 was checked in this manner. No oxygen content could be detected by the 9 μ absorption line.

---

[37] W. Kaiser, P. H. Keck, J. Appl. Phys. 28, 882 (1957).

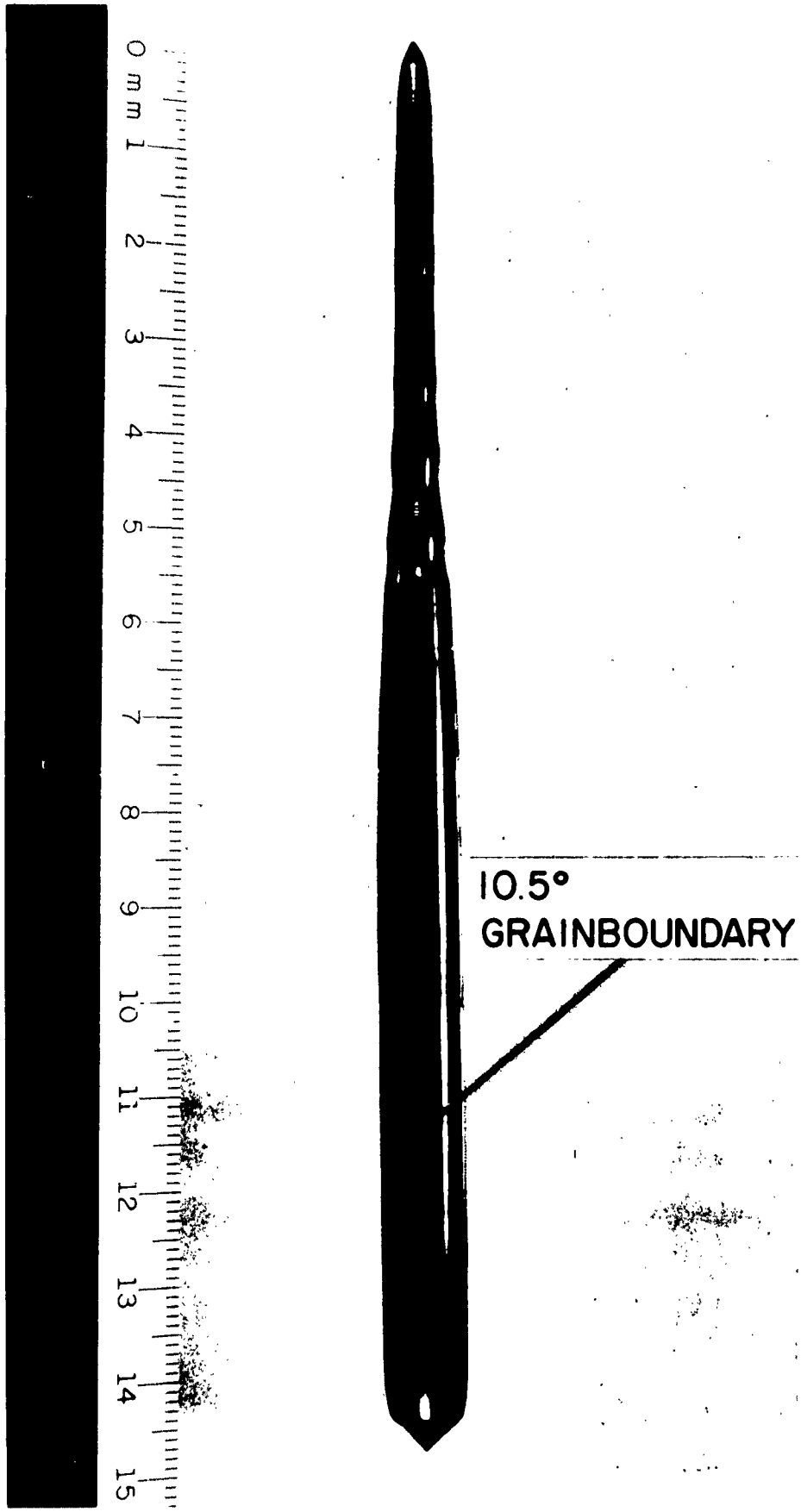


Figure 2.17

Oxygen free, 10.5° grain boundary crystal with [100] pull axis

D. Conclusions and Recommendations

Grain boundaries present an excellent means of studying dislocations and their effects. The studies described in this report lead to the following conclusions regarding the role of dislocations for semiconductor device reliability. Direct electrical effects of dislocations seem surprisingly weak. This speaks against the "dangling bond-model" in its simple form. P-n junctions are little affected by dislocations themselves. However, a considerable indirect influence has been demonstrated. The dislocations are surrounded by an atmosphere of impurities, which has been shown to be of great influence for electrical properites. Preferential formation of precipitates of foreign substances around dislocations can lead to serious failures of p-n junction devices. It is clear that such precipitates cause undesirable "soft" reverse characteristics. There is also evidence that the forward characteristic deteriorates if precipitates are present. Dislocations cause an enhancement of impurity diffusion. It is anticipated that this feature will eventually lead to utilization of grain boundaries, because by their unusual diffusion properties one can obtain structures which are impossible to construct with conventional means. The diffusion enhancement has consequences for device reliabilites. Very thin transistor base layers, for instance, may be effectively shorted out by such enhancements. Even if only a local reduction of width results,

serious failure of a device could be initiated by such "thin spots". Not enough quantitative information about these effects is known to date; further research seems necessary. The need for more sensitive techniques and even better defined material has become obvious from the studies done under this contract. With the trend towards smaller solid state components and higher power requirements it must be expected that failure mechanisms caused directly or indirectly by dislocations will become increasingly more important.*

* For a further, detailed discussion of these questions, the reader is referred to the Appendix. The possible influence of dislocations upon failure mechanisms in high power devices is treated in Secs. 7 to 10 of Part II in this report.

3. PROPERTIES OF TWIN BOUNDARIES IN SILICON

A. Introduction

Presented in this chapter are some investigations concerning twin boundaries and their influence upon structure and behavior of silicon devices. Twinning is a phenomenon very frequently observed in silicon. Several papers have dealt with the crystallographic structure of twin boundaries[38-43]. The most common boundary between two twins lies in a (111)-plane. Fig. 3.1 shows a model of this boundary, which is often called "first order" or" coherent" boundary. It is seen that the two grains are mirror images of each other. No dislocations, broken or distorted bonds are required for this boundary. It is obvious from Fig. 3.1 that the atoms at the boundary violate only second-nearest neighbor relations. Therefore, it is plausible that the energy to create this twin boundary must be low, hence the frequent occurrence of this twin boundary.

---

38 J. A. Kohn, Am. Mineral. 41, 778 (1956).

39 J. A. Kohn, Am. Mineral. 43, 263 (1958).

40 J. Hornstra, Physics 25, 409 (1959); 26, 198 (1960).

41 For general review on elastic and defect properties of Si, see P. Haasen and A. Seeger in Halbleiterprobleme, Vieweg. Braunschweig (1958) Vol. IV, p. 68, or

42 H. G. van Bueren, Imperfections in Crystals, North Holland Publishing Co., Amsterdam (1960).

43 A review on crystal growing and defects is given by E. Billig in Semiconductors and Phosphors, ed. by M. Schön and H. Welker, Interscience, New York and Fr. Vieweg, Braunschweig (1958) p. 2.

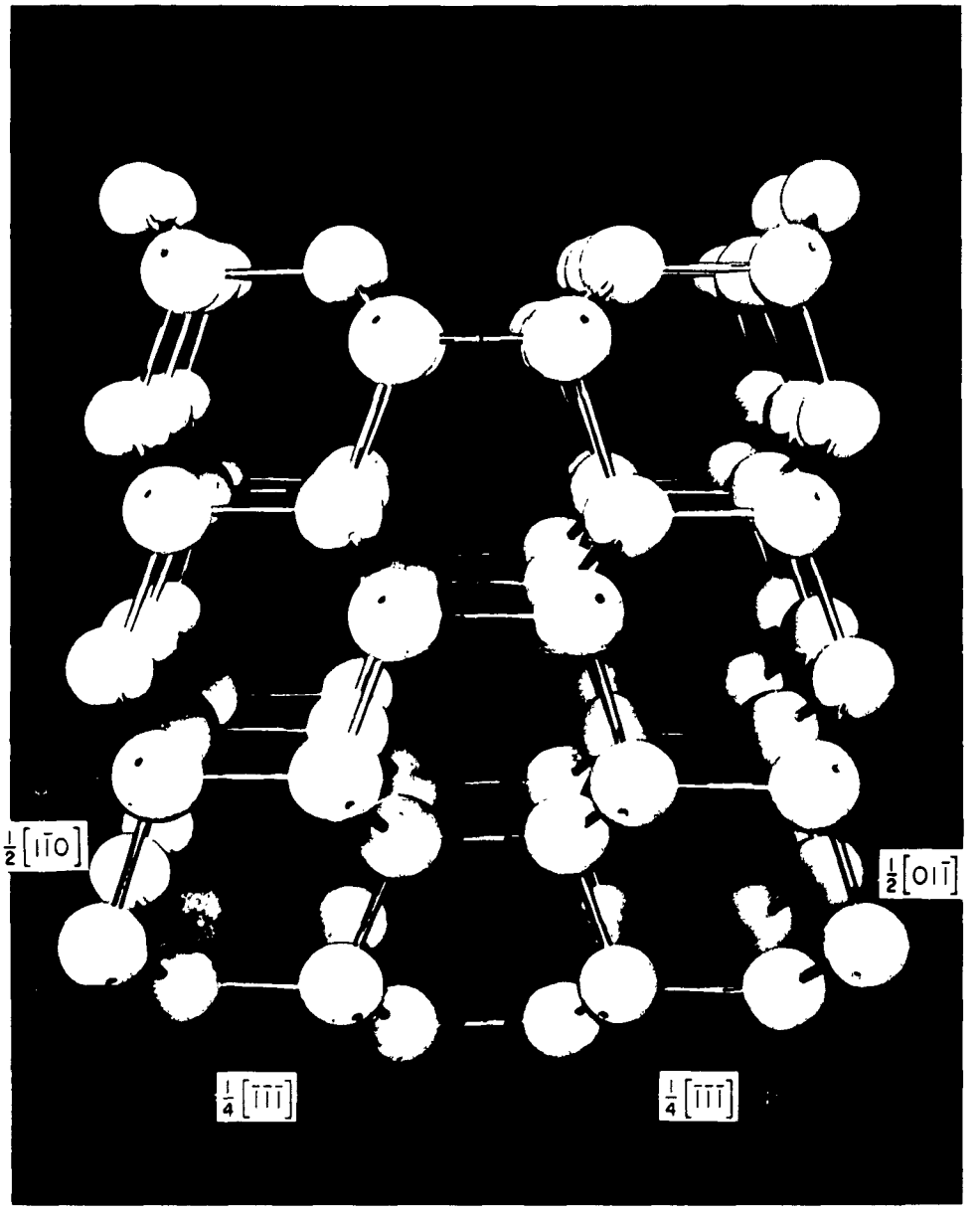


Figure 3.1

"Coherent," low-energy twin boundary in the diamond lattice. Boundary plane is mirror plane, it is (111)-plane for both grains. Common $[10\bar{1}]$-direction goes into paper.

Besides this low-energy fault, there are several other possible boundaries with higher energies, often generally categorized as "incoherent" or "high order" twin boundaries. Kohn[38, 39] and Hornstra[40] have discussed these boundaries in detail. One particular type of boundary will be of interest for our investigations. This structure is called "second order twin join" by Kohn[39]. Such a "join" is the interface between two grains ("individuals") which are in non-parallel first order (111) - twin relations to a third individual. Kohn shows that three such second order "joins" exist in the diamond lattice. He does not discuss the structure of these "joins" in terms of dislocations. Hornstra[40], on the other hand, maintains that all twin boundaries are only special cases of generalized grain boundaries, which are formed by regular arrays of dislocations.

This note reports various measurements made on an unusual, twinned silicon crystal. The specimen showed very regular boundaries; both the "coherent" (111) - boundary and a second order "join" were found within otherwise perfect crystalline material. Thus the crystal was particularly suitable for investigations of the two different kinds of boundaries. In the following, we shall first describe the structure of the sample and speculate about the causes of its peculiar growth. Measurements concerning electrical properties as well as precipitation and diffusion will be discussed next. Comparison with the structural models and some implications concerning p-n junction behavior at these boundaries conclude this paper.

B. Experimental

(1) Description of the Samples

A silicon single crystal was grown with a [100] pull axis, utilizing the Czochralski technique[43, 7] with a quartz crucible. Some uncontrolled interference during the early stages of the growth process resulted in simultaneous twinning on the four lower octrahedral (111)-planes. This gave four "coherent" twin-boundaries propagating towards the axis as the pulling proceeded.

Figure 3.2 (a) illustrates this situation. As the four boundaries moved inward, four new boundaries were created. These are the interfaces between two adjacent twins, each of which is in a "coherent" twin-relation to the original crystal. We thus have created* four "second-order joins," using Kohn's terminology[38, 39]. The coherent boundaries vanish after meeting at a common point on the crystal axis. The four high-energy joins are left "locked" into the crystal. Fig. 3.2 (b) further illustrates this development by showing the cross-sections of slices at different levels of the grown crystal. Fig. 3.3 shows photographs of lapped slices, corresponding to the drawings of Fig. 3.2 (b).

The growth of this unusual crystal was accidental. The unique character of the sample was not discovered until after slicing of the crystal boule. It may be speculated that a temperature shock or an

---

* This is an experimental verification of Kohn's idealized drawing; See Fig. 4 of Ref. 38.

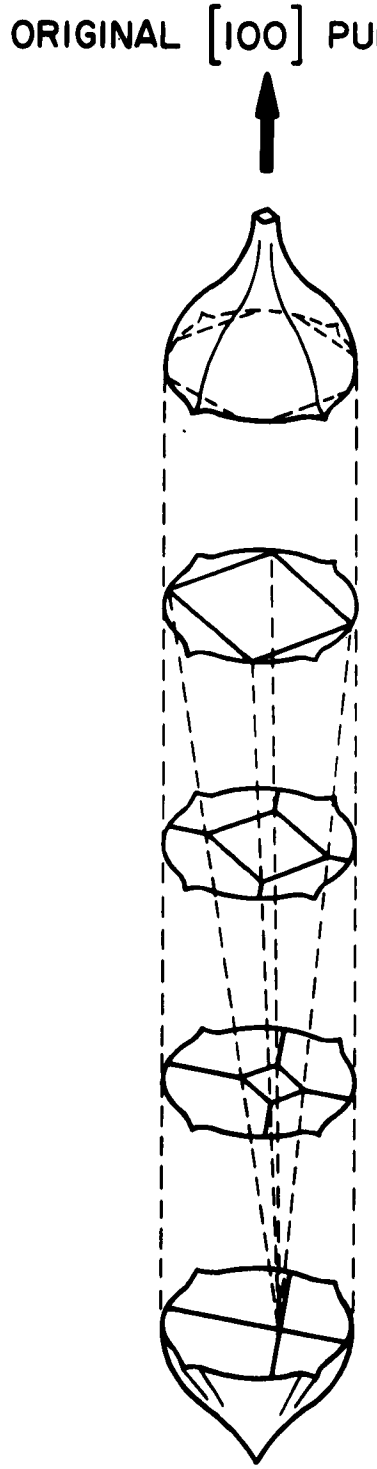


Figure 3.2a

Sketch of silicon crystal with grown-in boundaries

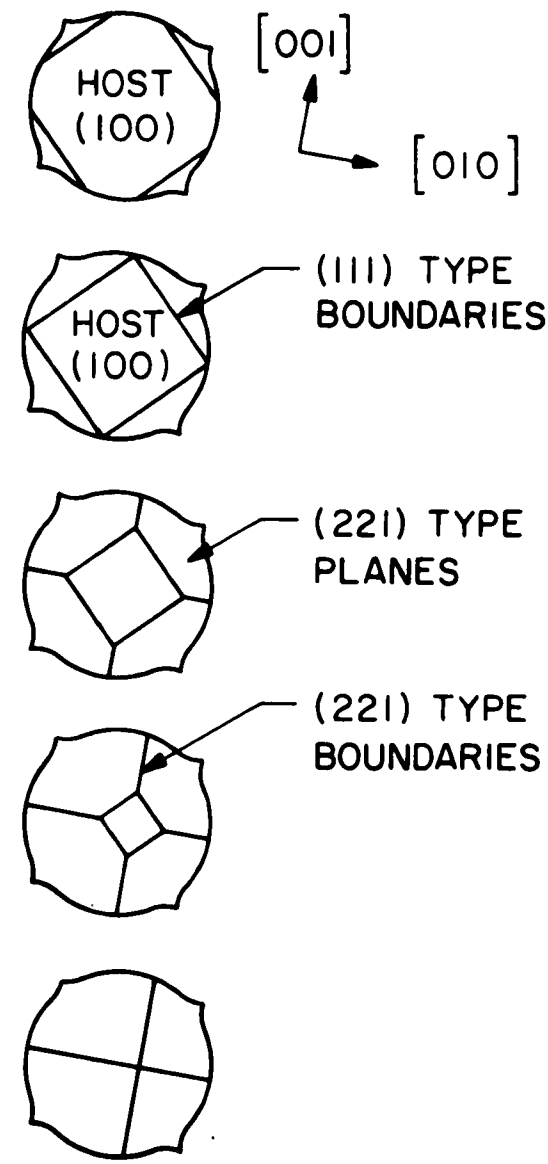


Figure 3.2b

Cross sections of crystal at different growth levels.

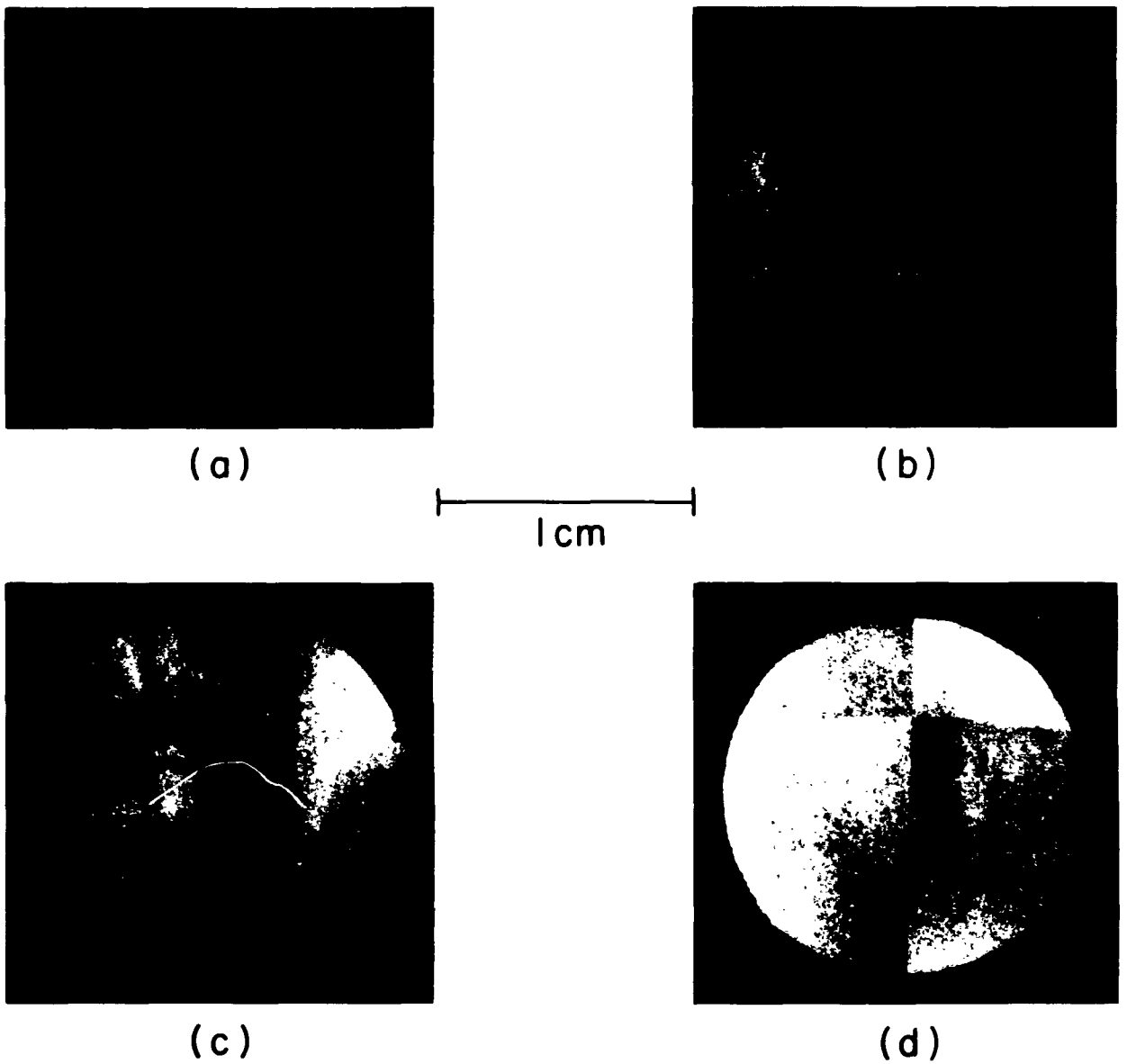


Figure 3.3

Photographs of crystal slices with twin boundaries, corresponding to drawing of Figure 3.3.

inadvertent sudden contamination of the melt led to the simultaneous twinning[43]. Such simultaneous twinning had frequently been found after deliberate introduction of "alloying mixtures"[44] into silicon melts. Although this seems a likely explanation, our data as well as those in the literature[44] are not sufficient to permit a conclusive statement.

The crystallographic evaluation of the slices was done first, roughly, by observing the slip patterns introduced by boron-diffusion[8]. A complete evaluation was then done with X-ray techniques*. It was established that the initial twinning occurred along (111) planes. The second-order "joins" were found to be of the $(2\bar{2}1)$-$(\bar{2}21)$ type. Both Kohn[38, 39] and Hornstra[40] give models for this type of twin boundary. Figure 3.4 is a reproduction of Kohn's proposal. It shows a zig-zag discontinuity with a resulting direction along $(2\bar{2}1)$ of twin individual A and $(\bar{2}21)$ of individual B. Arrays of nets with consecutively five and seven atoms are formed. No "dangling bonds" are necessary, but - as shown in the sketch - some bonds have to be distorted. Kohn does not speak of dislocations forming this boundary. Hornstra[40] derives

---

* We are very much indebted to Mr. William R. Cook of the Electronic Research Division, Clevite Corporation (Cleveland, Ohio) who kindly performed these investigations.

44 M. G. Mil'vidskii and L. V. Lainer, Sov. Phys., Solid State 3, 210 (1961).

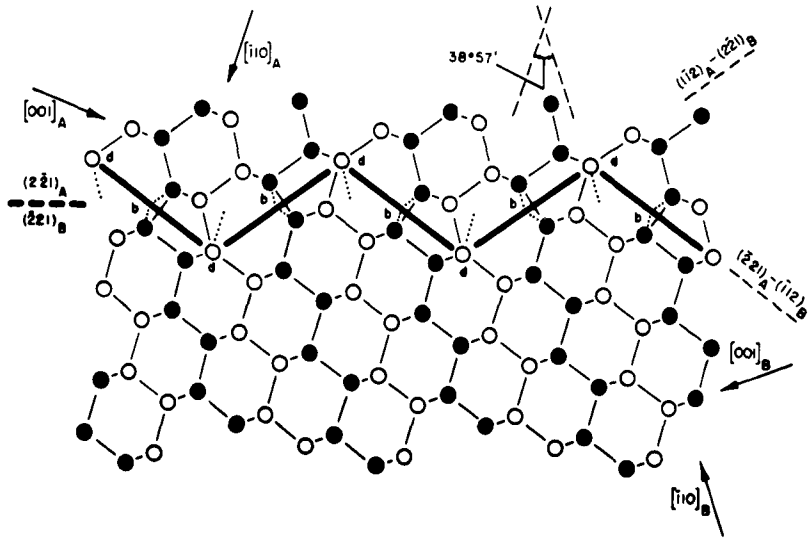


Figure 3.4

Structure of the (221)-type second order twin "join" after Kohn.[38,39] "A" and "B" refer to the two individuals of the twin. Bonds are distorted at points "d" and "b".

a structure identical to Fig. 3.4, but discusses it in terms of a closely spaced zig-zag array of dislocations. He presents an alternate model for this second order twin, which consists of overlapping double dislocations. This model is shown in Fig. 3.5.

(2) Electrical Properties

Slices from the crystal were cut perpendicular to the pull axis. They were then lapped and finally polished with "Linde A". The average electrical resistivity of the specimen was 1.5 ohm cm, p-type. The doping element was boron.

Photoscanning measurements[19] were done with a light spot of 25 micron diameter. Such measurements can give useful information about magnitude and type of the conductivity of boundaries in semiconductors. For example, germanium grain boundaries show photovoltages of a polarity which are thought to reveal the acceptor character of the "dangling" bonds in the boundaries. In our experiments, no photovoltage was measurable on either type of twin boundary (limit of detection about 10^{-6} volts). Conclusions from this result should be drawn with caution, since boundaries in silicon seem to show photo-effects different from those in germanium[32]. However, the absence of a photoresponse on the high energy boundary indicates that this boundary is probably not made up of dislocations with dangling bonds. The lack of a photoresponse on the "coherent" twin boundary is to be expected from its small lattice disturbance.

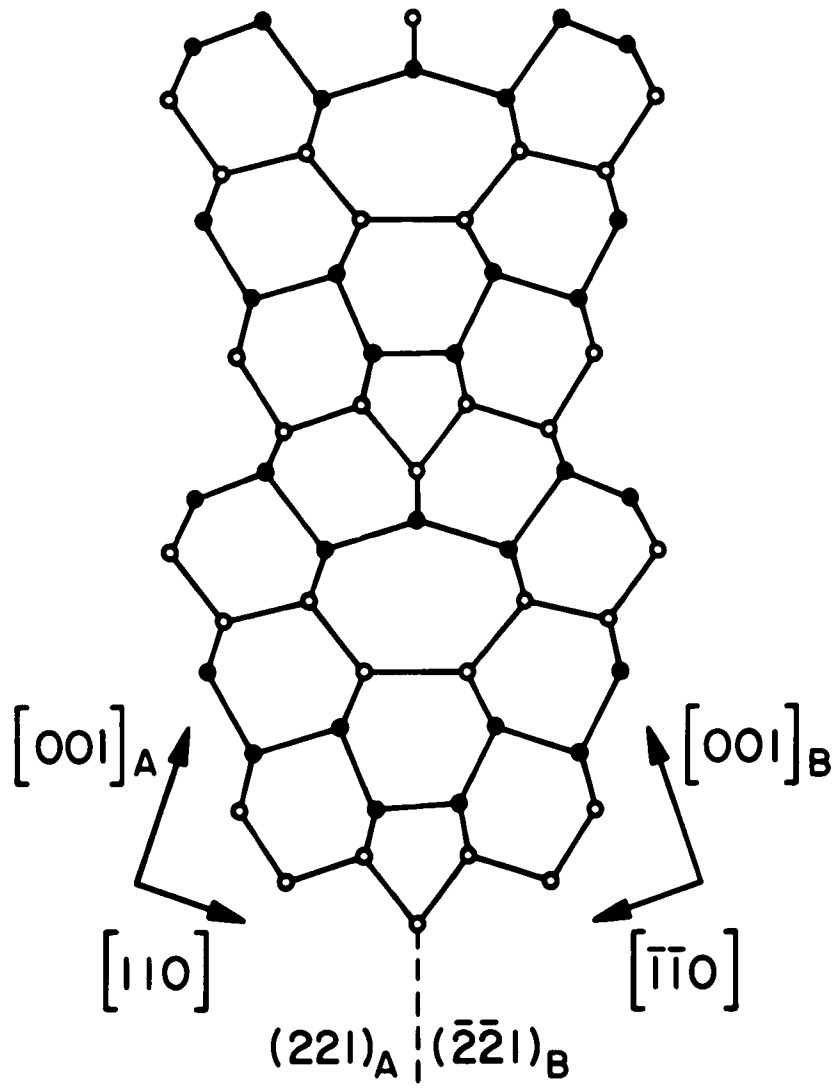


Figure 3.5

Alternate structure of (221)-type second-order twin boundary after Hornstra[40], assuming double dislocations

P-n junction diodes were produced by diffusing phosphorus from a P_2O_5 source into the slices at 950° C for one hour. Mesa diodes were etched by means of wax masking techniques. Care was taken to obtain as many mesas as possible that included the boundaries. Figure 3.6 shows photographs of the mesa diode arrays on two slices. The (221)-boundaries are perpendicular to the plane of the p-n junction, whereas the (111)-boundaries intersect the junction at an angle of approximately 55 .

The properties of the p-n junctions were investigated by measuring the reverse characteristics of the diodes. The mesas are classified as either regular or containing the coherent boundary or the second-order join. A distinction in junction behavior is made by calling a junction "soft" or "hard". "Softness"[14] means that before avalanche breakdown is reached, a comparatively large (in our case > 1 ma) reverse current is flowing through the diodes. This "softness" is an undesirable feature and usually constitutes a device failure mechanism.

Some typical results are shown schematically in Table 1.

Table 3.1

Evaluation of Diodes with and without Twin Boundaries

Reverse Current at 90% of breakdown voltage	< 1 μa "hard"	~ 1 μa	~ 10 μa	~ 100 μa	> 1 ma "soft"
Diodes on (221) boundary	30%	-	21%	10%	39%
Diodes on (111) boundary	100%	-	-	-	-
Regular Diodes	83%	3%	11%	-	3%

Figure 3.6

Etched mesa-diodes on slice with different twin boundaries.

It is seen that diodes on the "coherent" boundary were in all cases "hard", as were most of the devices on the regular material. On the other hand, the diodes which incorporated the high-energy boundary show a higher percentage of "softness". The reverse currents for these diodes were definitely higher than the average for the slice. From this set of measurements and similar ones on other slices we conclude the following. P-n junctions which incorporate a high energy twin boundary are very susceptible to "softness". Coherent boundaries do not have such a deleterious influence[45]. It is probable that the "softness" has been caused by metallic impurities in precipitated form[14]. Detailed measurements of soft reverse characteristics support this hypothesis. Figure 3.7 shows that the reverse current has a voltage dependence of approximately $V^{3.5}$. This is a dependence typical for metal precipitates[13]. "Gettering" with glassy oxide layers is known[14] to "harden" such soft devices. Some diffusion runs under these "gettering" conditions yielded hard diodes on the (221)-boundaries as well. This proves that the softness is not caused by the structure of the boundary itself. By copper decoration and infrared miscroscopy techniques[46] it was found that there is a tendency for preferential precipitation along the (221)-boundaries. No decoration was

---

45 E. Billig and M. S. Ridout, (Nature 173, 496 (1954) have also shown that coherent boundaries in germanium have no detrimental effects on electrical properties.

46 W. C. Dash, J. Appl. Phys. 27, 1193 (1956).

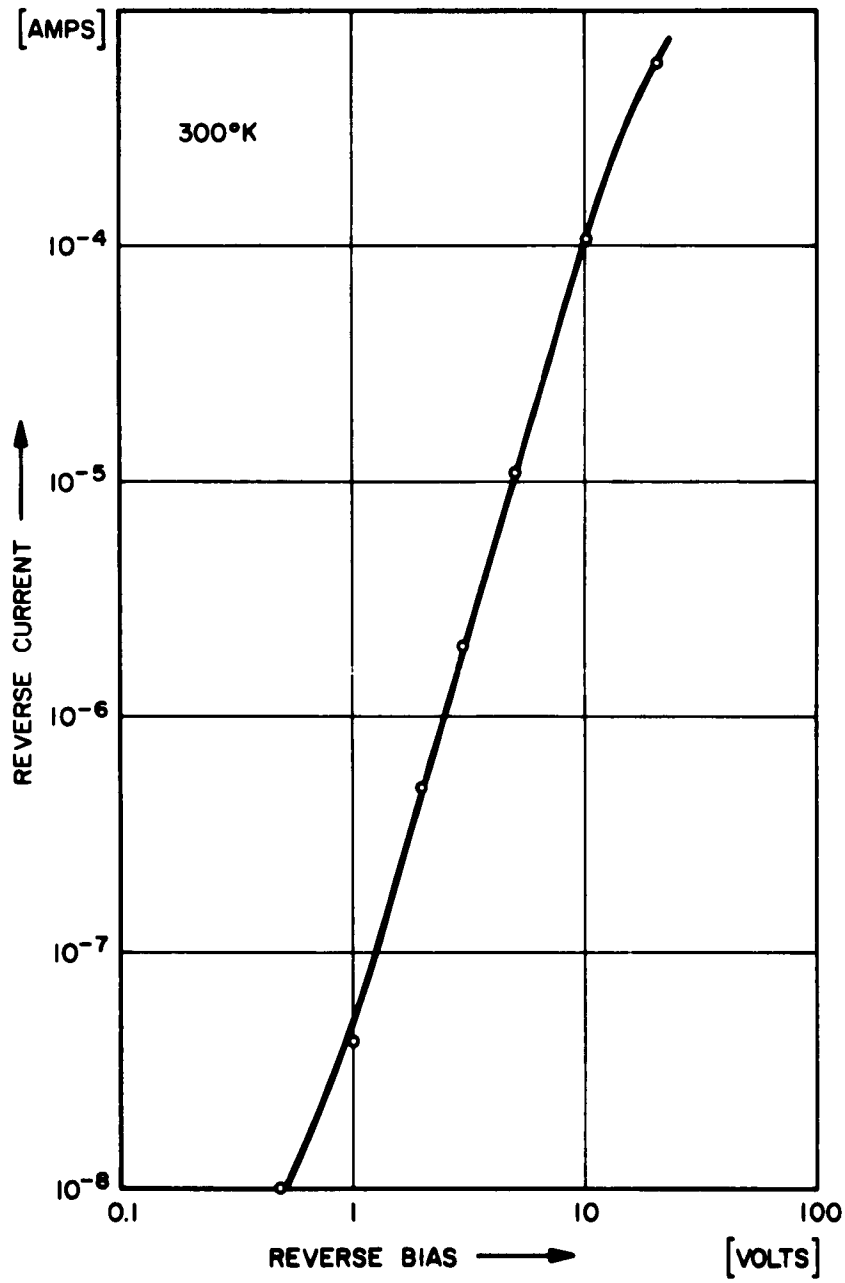


Figure 3.7

Reverse characteristic of a "soft" diode on a (221)-boundary, exhibiting the V^n-dependence of current upon voltage which is typical for metal precipitates

found on the (111) - twin boundary, in accordance with previous results[34]. It is thus very likely that the difference in precipitation behavior explains the different susceptibility for "softness" of p-n junctions.

The forward current-voltage characteristics of diodes on high energy boundaries show one peculiarity. These diodes obey a relation between current I and voltage V of the form

$$I = I_o' \, (\exp (qV/AkT - 1) \qquad (3.1)$$

where q is the electronic change, kT the Boltzmann constant times absolute temperature, and I_o' a fitting parameter. Theory predicts[15] the parameter A to take values between 1 and 2, but these diodes show A-values of about 3. Similar deviations in the forward characteristics have been found with silicon solar cells; one hypothesis offered to explain these anomalies on solar cells involved the assumption of precipitates[47]. The results on these twin boundary diodes thus seem to substantiate this hypothesis. Preliminary results, obtained in this laboratory with small angle grain boundaries also indicated high A-values in connection with precipitates.

---

[47] H. J. Queisser, Solid State Electronics 5, 1 (1962).

(3) Diffusion along Twin Boundaries

Coherent twin boundaries are reported to show no anomalies for impurity diffusion[48]. Several control experiments were performed with the silicon twin crystal described here. The techniques of previous grain-boundary investigations[2] were applied. No enhancement was found. Again, in view of the smallness of the disturbance which is introduced by the coherent boundary, this negative result is not surprising.

It is, however, not immediately obvious what kind of diffusion behavior should be anticipated at the high-energy second-order twin boundaries. If one considers these boundaries as being closely related to the artificially grown symmetrical grain boundaries, one should expect a diffusion enhancement similar to the one observed on the latter structures[2]. It is, however, important to consider the strain field built up by the boundary, since the Cottrell-attraction of the dislocations has been shown to be a decisive factor for the diffusion enhancement at grain boundaries[2]. Twin boundaries can be regarded as symmetrical grain boundaries with distinct angles of misfit. The grain boundary energy has minima at these angles[49]. This means that the strain field

---

48 S. Amelinckx and W. Dekeyser in Solid State Electronics, edited by F. Seitz and D. Turnbull, Academic Press, New York and London, (1959), Vol. 8, p. 468.

49 W. T. Read and W. Shockley, Phys. Rev. 78, 275 (1950).

must fall off more rapidly with distance from the boundary than for grain boundaries with a slightly differing angle of misfit. The "dislocation structure" of the twin boundary must be such that the strain fields of the very closely spaced dislocations almost cancel each other. In that case, one should expect a comparatively smaller diffusion enhancement, assuming that the Cottrell-attraction is the basic mechanism for the faster diffusion. In view of this unresolved situation it is of general interest to explore the structure of high-energy boundaries by making studies of impurity diffusion.

Several diffusion runs were performed with slices containing (221)-boundaries. A typical result is demonstrated in Fig. 3.8. This figure shows a beveled and stained[2] section of a slice which was diffused for one and one half hours at 1050° C with a P_2O_5 source. The p-n junction is indicated by the staining, the n-type diffused top layer appears light. The (221) twin-boundary is made visible by preferential etching, and appears as a dark vertical line. No diffusion "spike" [2] is observed. Thus there is no detectable diffusion enhancement.

Control runs were made with several other slices. By beveling at various angles possible diffusion enhancements in other crystallographic directions would have been detected. However, no such anisotropy was found; a "spike" was never observed. We thus conclude that under these diffusion conditions there is no detectable enhancement for substitutional impurity diffusion at the (221)-second-order twin boundary.

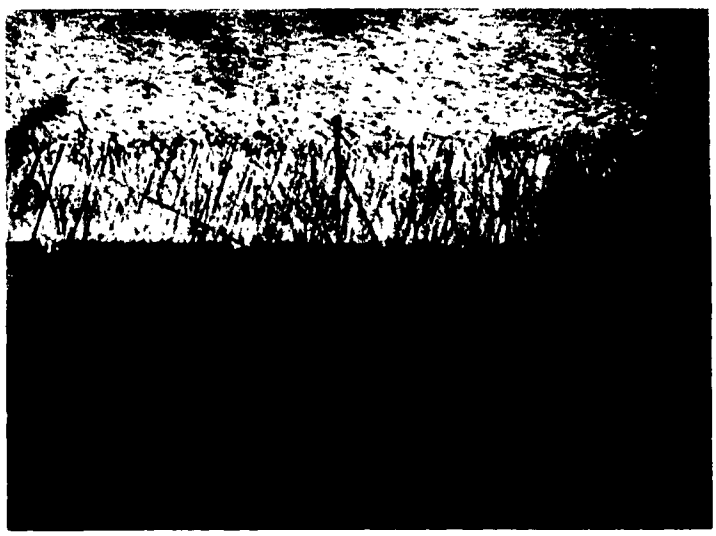


Figure 3.8

Diffusion front at (221)-twin boundary.
No evidence for diffusion enhancement.

C. Conclusions

It has been shown that coherent (111) twin boundaries in silicon have no unusual diffusion properties nor influence on p-n junction behavior. They appear indistinguishable from regular crystalline material.

High energy (221)-boundaries show a tendency for precipitation of impurities. This precipitation can induce "softness" in p-n junction diodes. The forward characteristics seem also to be influenced by precipitates. These twin boundaries do not enhance the diffusion of substitutional impurities. In this respect they differ strikingly from artificially grown grain boundaries which show strong diffusion enhancement.

The absence of photoresponse on either boundary type indicates that dangling bonds are unlikely. This is also substantiated by the fact that "hard" diodes are possible on uncontaminated boundaries which means that the grain boundary is not a good conductor. It can be summarized that the direct electrical effects on the "high-energy boundaries" are much weaker than expected[39, 41].

These results indicate that silicon devices should not be impaired by coherent boundaries. Similarly, only small direct effects should result from the stacking faults on (111)-planes, which have recently been detected in epitaxial silicon[50, 51], since stacking faults are

---

[50] H. J. Queisser, R. H. Finch and J. Washburn, J. Appl. Phys. 33, 1536 (1962).

[51] G. H. Schwuttke, J. Appl. Phys. 33, 1538 (1962).

closely related to twins[41-43]. The incoherent boundaries would be harmful for devices, but mostly indirectly through the action of impurity precipitates.

4. OBSERVATIONS OF LATTICE DEFECTS IN TRANSPARENT SHEETS OF SILICON

A. Introduction

During our work concerning defects in silicon we encountered the problem of fabricating thin silicon sheets suitable for electron transmission microscopy. The electron microscope work[52] will be discussed in the next chapter. Here we restrict ourselves to observations with a standard microscope. Monochromatic light is used and the specimens are observed in reflection. However, since the silicon samples are transparent, they could in principle also be examined in transmission with visible light.

B. Specimen Preparation

There are several possible methods to prepare ultrathin specimens for transmission studies. For instance, one can shoot a small jet of acid on to a sample[53, 54] until a hole is etched through. The edges usually have thin enough portions to be useful for electron microcopy. Electropolishing[55] has been tried successfully for similar substances. Silicon has also been thinned down by cathode sputtering[56] until it became transparent.

---

52 H. J. Queisser, R. H. Finch and J. Washburn, J. Appl. Phys. 33, 1536 (1962).

53 G. R. Booker and R. Stickler, El. Chem. Soc. Meeting Los Angeles, May 1962, Abstract 71.

54 B. A. Irving, Brit. J. Appl. Phys. 12, 92 (1961).

55 R. P. Riesz and C. G. Bjorling, Rev. Sci. Instr. 32, 889 (1961).

56 B. Hietel and K. Meyerhoff, Z. Physik 165, 47 (1962).

A chemical method is applied in our investigations[57]. Constant agitation of the sample insures uniform etching as follows. A tilted rotating beaker contains the etching solution, consisting of 95% HNO_3 and 5% HF (both concentrated). The sample is a silicon slice; it is fastened to a round teflon disk with apiezon "black wax". The side to be investigated is placed towards the disk. The tilted beaker is rotated at 10 to 30 rpm and the disk is constantly rolling around inside the beaker. This insures constant agitation and uniform etching.

Almost the entire slice is etched off from the reverse side until there is only a thin layer left. The etching is done until at some spot a hole appears and the "black wax" substrate becomes visible. The edges around such a hole are usually very thin wedges with an extremely small wedge angle. We thus obtain sufficiently large areas of a thickness which easily permit observation in transmission.

C. Observations

(1) Defects in Epitaxial Layers

The most convenient way of observation is to leave the thin layer on the supporting wax substrate and to inspect the sample in reflection with monochromatic light. A sodium vapor lamp was used for our studies.

---

57 M. V. Sullivan and R. M. Finne, paper presented at the Houston Meeting of the Electrochemical Society, October 1960.

Several types of silicon were examined. Figure 4.1 gives a typical result for epitaxial material. This slice was grown by decomposition of $SiCl_4$. The black substrate is seen in Figure 4.1. Interference fringes are observed on the silicon surfaces. These fringes are contours of equal thickness. They are caused by reflection on top and bottom of the silicon wedge. It should be pointed out that no fringe plate is used as would be the case in a multiple beam interference experiment. One spacing between adjacent fringes corresponds to a thickness change of only about 700 Å, because the index of refraction is approximately 3.4 in this wavelength region. Figure 4.1 shows rather uniform sections but also typical lattice defects, like growth pyramids and dislocations. These defects are frequently found in slices that were epitaxially grown. Also some triangular structures can be detected in Figure 4.1. These shall be discussed in some more detail.

Figure 4.2 shows another view of a (111)-plane of epitaxial silicon. Triangles and short ribbons are seen. Electron microscopy[52] has identified these structures as stacking faults. The closed triangles are the bases of tetrahedrons of (111)-planes with the vertices at the interface between substrate and epitaxial deposit. Figure 4.3 shows these faults in a higher magnification. The tetrahedrons grow in size with increasing film thickness. Figure 4.3 shows clearly that these tetrahedrons have intergrown.

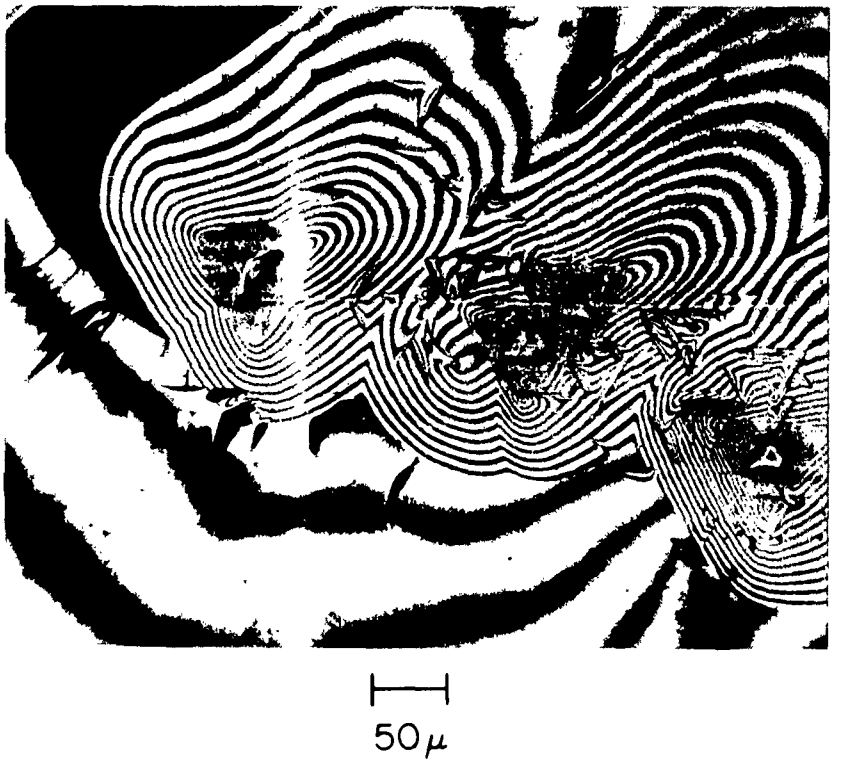


Figure 4.1

Epitaxial silicon after etching. Black area is wax substrate, interference fringes of sodium light give contours of equal thickness. Defects typical for epitaxy (growth pyramids and stacking faults) are shown.

Figure 4.2

Stacking faults on a (111) - face of epitaxial silicon

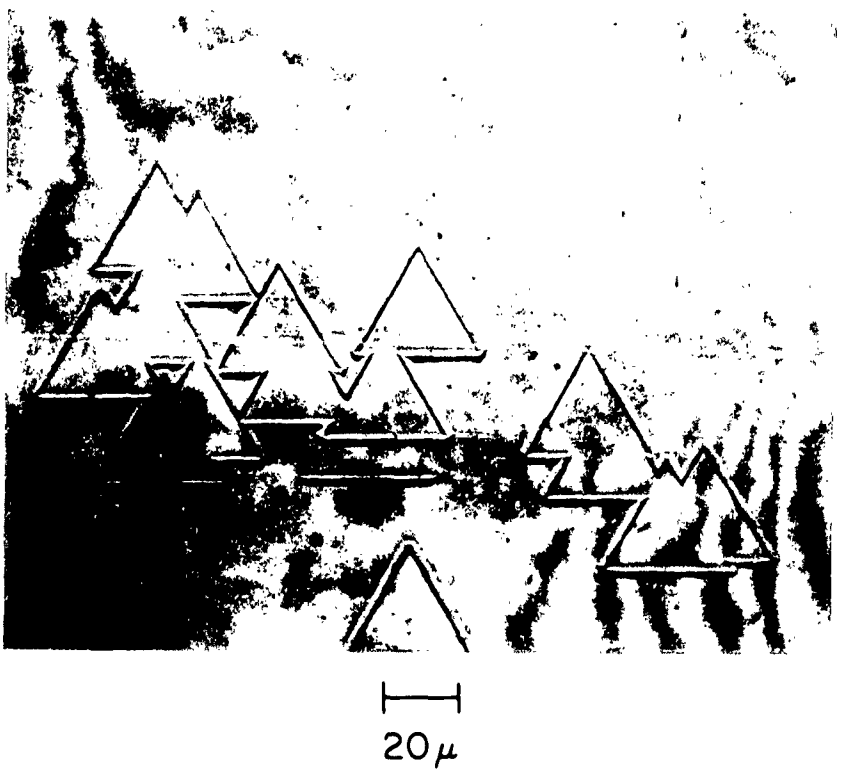


Figure 4.3

Enlargement of a section of Fig. 4.2

Figure 4.4 shows stacking faults on a film which was grown on a (100) substrate. We now see the (111) tetrahedrons in a different projection. They appear to be nearly squares. From the figure it is seen that the etching was quite uneven. Surprisingly, one even finds replicas of the stacking faults in the wax substrate. Apparently, the etch attacked the wax where the silicon was etched through first. One can see that the areas with large concentrations of faults etched more rapidly than the rest of the silicon.

It is remarkable that the stacking faults etch as preferentially as is shown in these figures. It may be speculated that there is also some sort of impurity precipitation encountered at these faults, which accelerates the preferential etching.

(2) Diffusion-induced Dislocations

Another problem which is of interest concerns diffusion-induced dislocations[8]. If high concentrations of undersized substitutional impurities-such as boron-are diffused into silicon, stress will be introduced which is relieved by the formation of dislocations. This has been shown with several methods[8, 58, 59] but insufficient information is thus far available about the distribution and penetration depth of these dislocations. It is possible that with the aid of the described thin sectioning, information can be obtained about the penetration depth of the diffusion induced dislocations.

---

[58] S. Prussin, J. Appl. Phys. 32, 1876 (1961)

[59] G. H. Schwuttke and H. J. Queisser, J. Appl. Phys. 33, 1540 (1962).

Figure 4.4

Stacking faults in an epitaxial film grown on a (100) substrate.

Figure 4.5 shows boron diffused sample after etching and a brief treatment with Dash etch[46]. Besides the interference fringes, one can observe the diffusion induced dislocations which have been brought out by the Dash etch treatment. The distorted fringe lines make the dislocations clearly visible. In principle one can count the dislocations and relate them to the thickness which is measurable by means of the interference fringes. Preliminary results seem to indicate that the dislocations penetrate much deeper than a diffusion length of the boron impurities. There also seem to be two regions of high dislocation concentrations, one close to the surface and another one at a larger depth. In between there seems to be a region with a lower dislocation density. At the moment this phenomenon is not understood and further experiments should be made to verify these first observations.

(3) P-N Junction Delineation

Finally, there is another phenomenon which could become a useful tool for investigations of shallow diffused layers. Figure 4.6 shows another boron diffused slice after etching and Dash etch treatment. A rather pronounced dark line is seen. At the moment it is not quite clear what causes this dark line, but it seems to be a region of strongly enhanced attack by the Dash etch. Correlation with other measurements indicates that the depth is the same as that of the diffused p-n junction. Thus the black line separates the diffused p layer from the

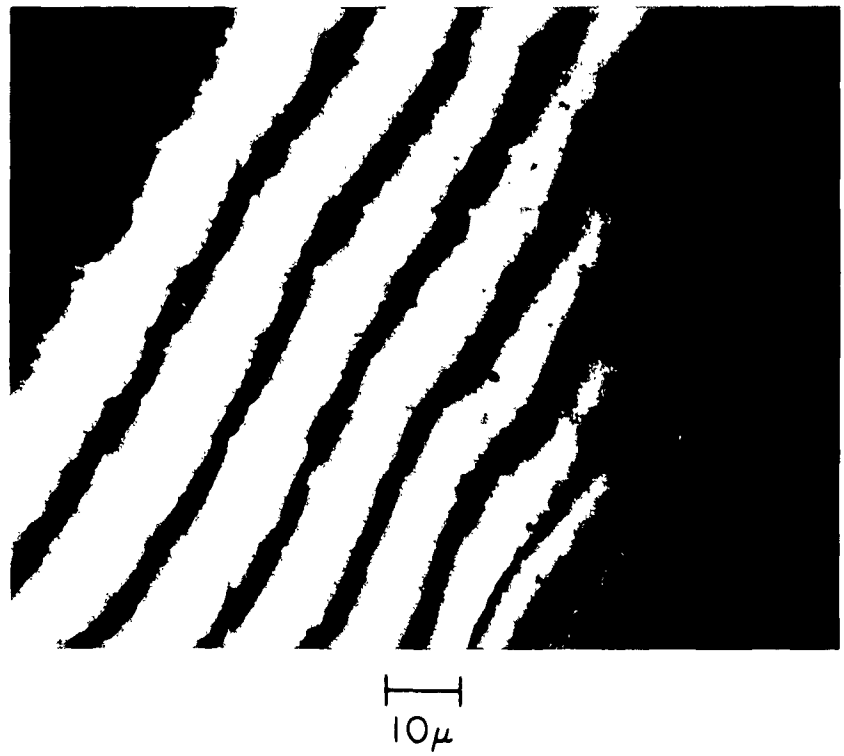


Figure 4.5

Thin wedge of a silicon slice with diffused boron layer. Diffusion induced dislocations form rectangular network.

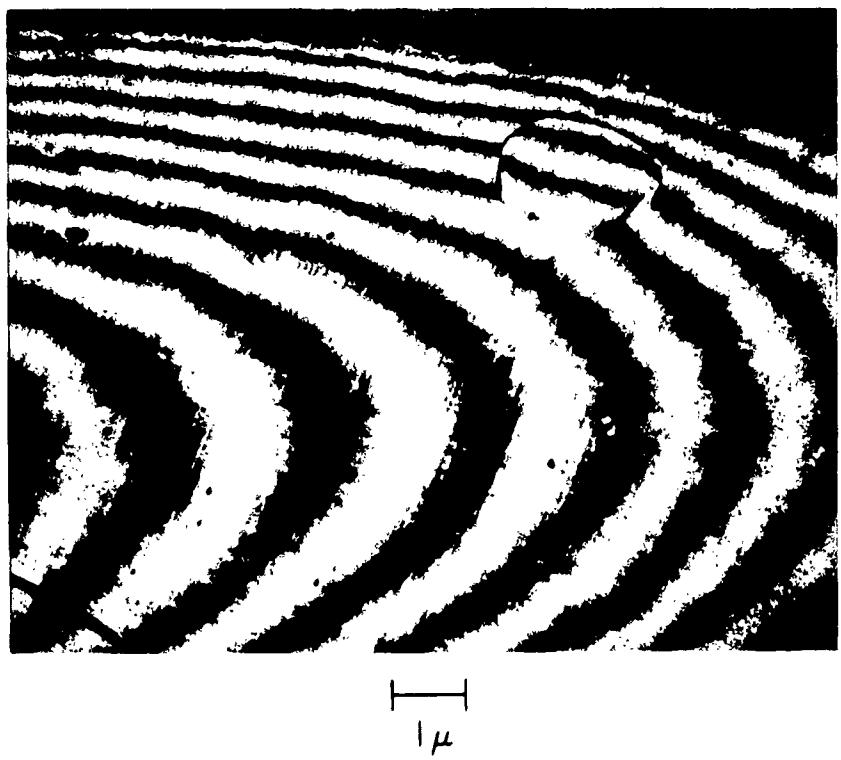


Figure 4.6

Boron-diffused slice after thinning and Dash-etch treatment. Heavy black line in upper part delineates p-n junction. Semicircular irregularity in the upper right portion is unidentified region of accelerated etch attack, possibly caused by doping variation.

base n material. It is also seen that the etch attacks these two layers differently. This is seen by the difference of the spacing of the interference fringes. This method of delineating p-n junction seems to work only with boron or phosphorus diffusions. Gallium diffusions did not show this effect. One could therefore conclude that dislocations are of importance for this junction delineation, since Ga-diffusions do not introduce dislocations[59].

It is seen that this method could be useful for the determination of junction depth of the very shallow diffused layers. One simply has to count the number of fringes towards the black line. Because of the small thickness difference that corresponds to one fringe spacing, the method is accurate but destructive.

5. STRUCTURE AND ORIGIN OF STACKING FAULTS IN EPITAXIAL SILICON*

A. Introduction

Epitaxial films of silicon can be grown upon silicon-single crystal substrates by thermal decomposition of volatile compounds, such as $SiCl_4$[60]. Several investigations have revealed crystalline defects which are typical for such epitaxial material. Light[61] and Haase[62] reported triangular defects which were observed on (111)-planes of epitaxial silicon and germanium. These authors inferred that each triangle corresponed to a tetrahedron with one corner at the interface between substrate and film. It was shown later by electron microscopy[50] and x-ray investigations[51,63] that these defects are stacking faults. Stacking faults have been shown to exist as either closed figures, or as ribbons bounded by dislocations having a fault vector a/6 <112> [63].

B. Experimental

For transmission electron microscopy, samples of epitaxial silicon < 1μ thick are needed. Such samples may be prepared by first growing a

---

* This work was partly supported by the Inorganic Materials Research Division of the Lawrence Radiation Laboratory, U. S. Atomic Energy Commission (Professors J. Washburn and G. Thomas).

60 H. C. Theuerer, J. Electrochem. Soc. 108, 649 (1961).

61 T. B. Light, AIME Conf. Metallurgy of Semiconductor Materials 1961, Interscience, New York, (to be published).

62 O. Haase, AIME Conf. (See Ref. 2).

63 G. H. Schwuttke, Meeting Electrochem. Soc., Los Angeles, May 1962.

relatively thick layer by vapor decomposition and subsequent etching to the desired thickness.

Epitaxial silicon samples were grown by the commonly used thermal decomposition of $SiCl_4$ with hydrogen[60]. Substrate temperatures range from 1070° C to 1125°C, as measured with an optical pyrometer; the actual temperatures were estimated to be higher by approximately 50°C to 60°C. Single crystalline substrate slices were used with various crystallographic orientations of the surfaces. The slices were placed on a silicon coated graphite pedestal inside a quartz reaction chamber and heated by induction. A typical growing time was 20 minutes, resulting in a thickness of 10 to 15 microns for the epitaxial layer. Substrate resistivities were of the order of 0.5 ohm-cm, the layer resistivities averaged 1.0 ohm-cm. Undoped $SiCl_4$ was used.

The samples grown in this manner were etched by a method similar to the one described by Sullivan and Finne[56]. The silicon slice was mounted with wax onto a teflon disk of approximately 2.5 cm diameter, with the epitaxial layer downward. The disk was placed inside a plastic beaker, which was tilted at an angle of 45° and rotated at approximately 30 rpm. The sample rolled constantly around as the beaker rotated. This agitation ensured uniform etching. The polishing solution consisted of 95% HNO_3 and 5% HF (both concentrated).

Etching was continued until at one spot the sample was removed entirely so that the wax substrate became visible. Optical examination

revealed the sample thickness through interference fringes around the edges of such holes, thus enabling specimens to be obtained which were thin enough for electron transmission.

The optical observations[64] were made while the sample was still on its teflon support. It was convenient to investigate the specimen with monochromatic light in reflection. Interference fringes were observed, which were caused by reflection at the top and bottom surfaces of the foil. From these fringes one can estimate the foil thickness and judge the uniformity of the etching. Lattice defects cause irregular etching and become clearly visible under the optical microscope, especially with the aid of the interference contours. Replicas of etch patterns from lattice defects were also seen on the wax substrate, since the wax was also attacked by the polishing solution, thus revealing the spots where the silicon was removed first.

Samples with a sufficiently high density of faults were selected for further observations with the electron microscope. The foils were detached from the teflon disk by dissolving the wax in an organic solvent. Suitable pieces of sufficiently thin sections could be easily broken off and mounted in the sample holder of the electron miscoscope. Transmission electron microscopy was done using a Siemens Elmiskop 1A microscope operated at 100kv.

---

64 H. J. Queisser and R. H. Finch, Bull. Am. Phys. Soc. 7, 211(1962).

C. Results

(1) Structure for Various Substrate Planes

Triangular closed figures of stacking faults were observed on epitaxial films grown upon a substrate with a (111)-surface. Figure 5.1a gives an example, showing several triangular defects, having grown to such a size that they intersect each other. The stacking faults follow <110> directions. The triangles represent the intersection of the surface with a tetrahedron consisting of {111} -planes.

An electron micrograph typical of stacking faults in layers grown on (111) substrates is shown in Fig. 5.1b. Here four tetrahedral defects have merged together. The existence of a fault between AB shows that each side of the tetrahedron is composed of two or more close, parallel, faults. The similarity between Figs. 5.1a and 5.1b is quite striking.

Equilateral stacking fault triangles were observed when the substrate surface was a (110)-plane. (Fig. 5.2a) It is seen that different portions of the triangles show stronger etching and wider grooves than other portions. It is possible that more rapid attack of the etchant may occur where there are several close parallel faults. The electron micrographs, reproduced in Figs. 5.2b and 5.2c, show clearly that faults at 90° to the surface also etch much more rapidly than those on the 35° planes.

Experiments with substrate surface near (100) yielded square-shaped stacking fault figures such as the ones shown in Fig. 5.3a. The diagonals in these squares are also revealed. This may be due to the

Figure 5. 1a

Optical micrograph showing closed figures of intergrown stacking faults grown on substrate with (111)-surface. Observation in reflection with monochromatic light (Na-lamp). Interference contours caused by reflection on top and bottom surface of transparent foil.

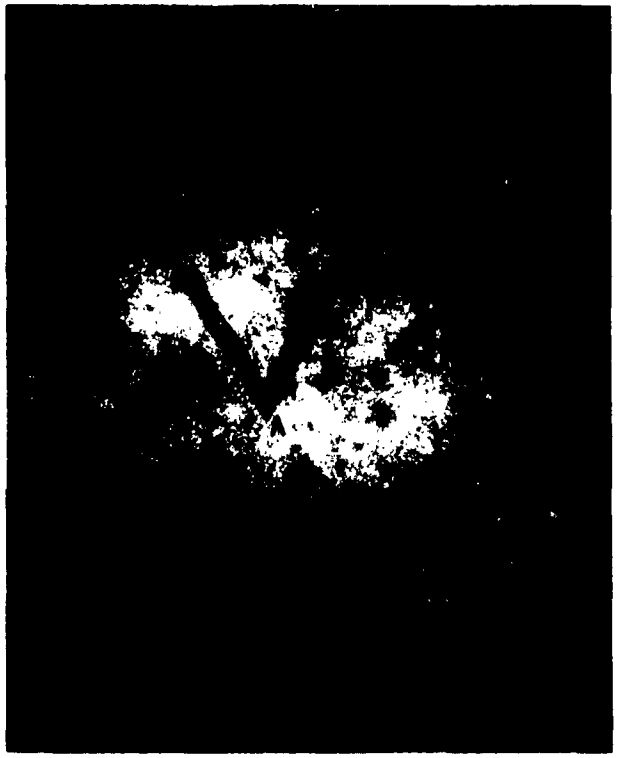


Figure 5. 1b

Electron micrograph of silicon crystal prepared from slice similar to that in Fig. 5. 1a. The change in contrast at AB shows there are many overlapping faults present. [111] foil.

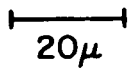


Figure 5.2a

Optical micrograph showing unevenly etched stacking faults grown on (110)-substrate.

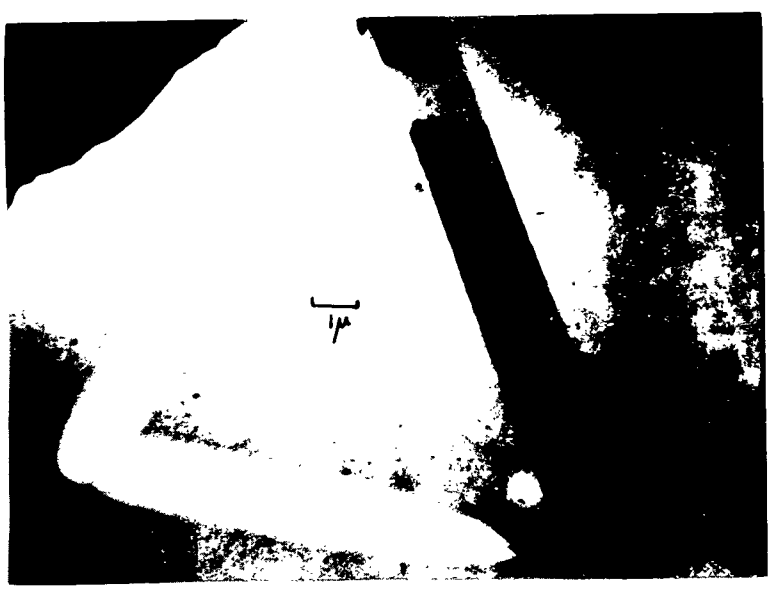


Figure 5.2b

Electron micrograph corresponding to Fig. 5.2a. Notice preferential etching of {111} normal to foil surface. [110] foil.

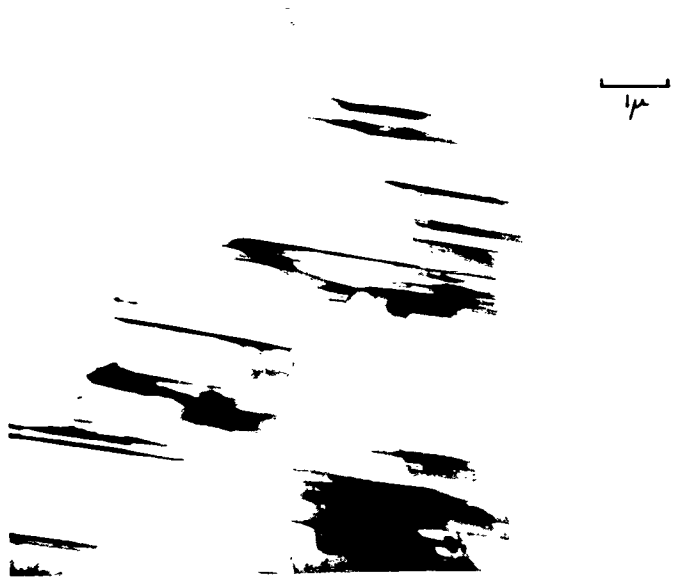


Figure 5.2c

As 5.2b. Showing dense region of faulting and preferential etching.

effect of the faults on the contour of the growing surface. A typical corresponding electron micrograph is shown in Fig. 5.3b. In Fig. 5.3b the non-parallelism of the faults at AB and those adjacent is due to the specimen surface being about 7° off from exact (100) orientation. It is also clear from this micrograph that bundles of dislocations e.g., at C. sometimes were observed near the ends of long faults.

It can be concluded from these observations that stacking faults in (111)-planes occur as either closed figures or dislocation-bounded ribbons, regardless of the orientation of the substrate. The optical observations indicate that the tetrahedral closed figures seem to be most frequent, but the electron micrographs demonstrate that short ribbons or very small closed figures may occur which cannot be detected optically.

(2) Identification of Faults

The kinematical theory of phase contrast can be used to identify faults and their bounding dislocations. In this theory, Hirsch et al[65] have shown that dislocations are in contrast only when $\bar{g}.\bar{b} \neq 0$ i.e., when the Burgers vector $\bar{b}$ is not parallel to the reflecting plane (hkl) which is giving rise to the reciprocal lattice vector $\bar{g}$. This criterion

---

65 P. B. Hirsch, A. Howie and M. J. Whelan, Phil. Trans. Roy. Soc., A, 252, 499 (1960).

Figure 5.3a

Stacking faults in Si grown on substrate with (100)-surface.

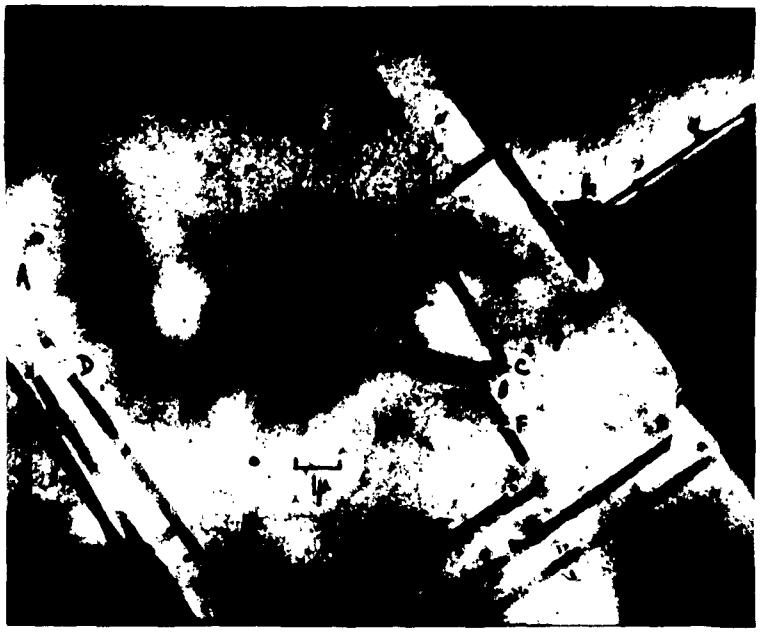


Figure 5.3b

Electron micrograph of silicon crystal in [100].

is also true in the dynamical theory. For stacking faults a similar criterion for contrast exists[66]. In this case contrast is obtained when the phase shift $2\pi\bar{g}.\bar{R}$ due to faulting is $\neq 0$ or $n2\pi$ ($\bar{R}$ is the displacement vector for faulting). Values of the quantities $\bar{g}.\bar{b}$ for partial dislocations with $\bar{b}$ = a/6 <112>, a/3 <111> and a/6 <110> are given in Table 5.1 for the common strong reflections obtained from the oriented crystals.

This table shows that of the principal reflections, 220, 400, from foils in [100] or [111] orientation, only two of the possible four faults should be in contrast for the $2\bar{2}0$ reflection, whereas they should all be visible for the $0\bar{4}0$ reflection. Figure 5.4a is a micrograph of a foil in [100] with a strong $0\bar{4}0$ reflection operating. Both faults on (111) and $(1\bar{1}1)$ are visible as well as the bounding dislocations. Upon tilting to utilize the $2\bar{2}0$ reflection, in accordance with Table 5.1 the (111) fault disappears whereas the $(1\bar{1}1)$ fault is still visible. However the bounding dislocation for the (111) fault is still visible. Examination of Table 1 shows that the Burgers vector cannot be a/6 $[\bar{1}\bar{1}2]$ (or equivalent ones for this plane) or a/3 [111] since these should be invisible - the only other possibility is an a/6 <110> Burgers vector i.e., a stair-rod dislocation. From the geometry of Fig. 1 this Burgers vector must be

---

66 M. J. Whelan and P. B. Hirsch, Phil. Mag. 2, 1121, 1303, (1957).

Table 5.1

Values of the quantities $\bar{g}.\bar{b}$ for partial dislocations and for principal reflections obtained from Si foils in [100], [110] and [111] orientations

Fault Plane	$\bar{b}$	$\bar{g}.\bar{b}$ 400	040	220	220	111	1$\bar{1}\bar{1}$	4$\bar{2}\bar{2}$	311
(111)	a/6[$\bar{1}\bar{1}$2]	-2/3	2/3	0	-2/3	0	-1/3	0	-1/3
	a/6[2$\bar{1}\bar{1}$]	-2/3	2/3	0	1/3	0	2/3	0	2/3
	a/6[$\bar{1}$2$\bar{1}$]	-2/3	2/3	0	1/3	0	-1/3	0	-1/3
(11$\bar{1}$)	a/6[121]	-2/3	-2/3	1/3	0	-1/3	0	-1/3	-2/3
(1$\bar{1}\bar{1}$)	a/6[$\bar{2}\bar{1}$1]	2/3	2/3	-1/3	0	-1/3	-1/3	2/3	0
($\bar{1}$11)	a/6[$\bar{2}$1$\bar{1}$]	2/3	-2/3	0	-1/3	-1/3	-1/3	2/3	0
(111)	a/3[111]	4/3	-4/3	0	4/3	1	-1/3	0	5/3
(11$\bar{1}$)	a/3[11$\bar{1}$]	4/3	-4/3	0	4/3	1/3	1/3	4/3	1
(1$\bar{1}$1)	a/3[1$\bar{1}$1]	4/3	4/3	4/3	0	1/3	1/3	4/3	1
($\bar{1}$11)	a/3[$\bar{1}$11]	-4/3	-4/3	-4/3	0	1/3	-1	-2/3	-1/3
(111)	a/6[110]	2/3	-2/3	0	2/3	1/3	0	1/3	2/3
	a/6[011]	0	-2/3	-1/3	1/3	1/3	-1/3	-2/3	1/3
	a/6[101]	2/3	0	1/3	1/3	1/3	0	1/3	2/3
(1$\bar{1}$1)	a/6[$\bar{1}$0$\bar{1}$]	-2/3	0	-1/3	-1/3	-1/3	0	-1/3	-2/3
	a/6[1$\bar{1}$0]	2/3	2/3	2/3	0	0	1/3	1	1/3
	a/6[0$\bar{1}$1]	0	2/3	1/3	-1/3	0	0	0	0

a/6 [101]. This dislocation should be in contrast both for the 400 and $2\bar{2}0$ reflections (Table 1) as is actually found. Both a/6 <112> and a/6 <110> dislocations were detected by Schwuttke using x-ray techniques[51,63].

Our results show that the contrast at A in Fig. 5.4 is due to two parallel stair-rod dislocations. This configuration can be explained if the fault on (111) bends over along $(1\bar{1}1)$ and returns parallel to the original (111) fault. Thus what appears to be a simple fault system is in fact a closed figure. Consideration of this figure shows that if one of the (111) faults is intrinsic, the parallel fault must be extrinsic, as sketched in Fig. 5.5.

The only other explanation of the contrast that we can think of is that the dislocations at A in Fig. 5.4 are made visible by some complex multiple diffraction and/or absorption phenomena. However since the fault contrast obeys the simple phase-contrast rule one expects the dislocations to do so also, so that this latter possibility is considered to be an unlikely one. This is the first observation of pairs of intrinsic and extrinsic faults and a discussion of a possible mode of nucleation is left until later (section D).

At first sight we might expect that overlapping intrinsic-extrinsic faults would give rise to zero contrast since the phase shifts are equal but opposite in sign. Consideration of the contrast from kinematical diffraction theory shows that for two overlapping faults of opposite sign there will

Figure 5.4a

[100] Foil, strong $0\bar{4}0$ reflection operating.

Figure 5.4b

After tilting, $2\bar{2}0$ reflection operating: notice disappearance of fault along (111) but dislocations A remain in contrast.

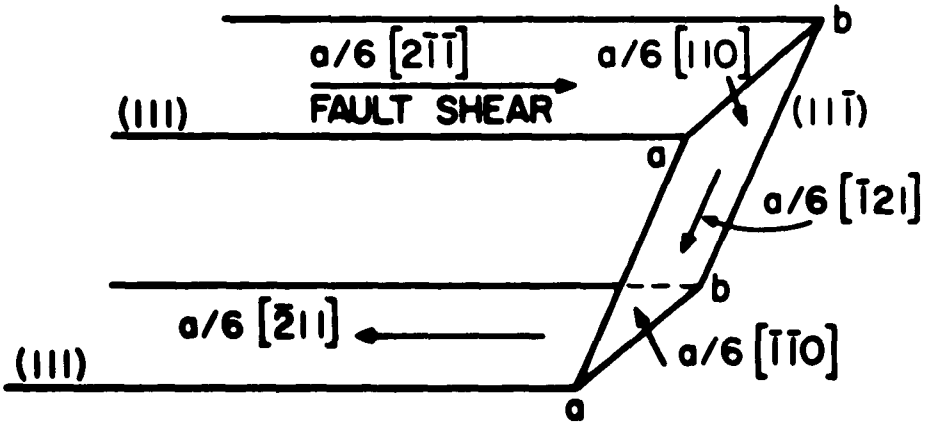


Figure 5.5

Schematic sketch showing the nature of the intrinsic-extrinsic stacking fault figures suggested from Fig. 5.4. It is the likely configuration along the sides of many of the tetrahedra of Figs. 5.1 - 5.4 ab are the connecting stair-rods (see also Fig. 5-13).

be zero contrast only when the two faults are very close together or when the faults are separated by an integral number of extinction distances. When the separation is not too small a fraction of an extinction distance, overlapping contrast is expected. The contrast then becomes similar to that for overlapping faults of the same sign. The separation of the two dislocations A in Fig. 5.4b is $\sim$ 200 A i.e., $\sim$ one -quarter of the extinction distance for the $2\bar{2}0$ reflection (see Table 5.2), hence we believe that our conclusions regarding the fault contrast are not unreasonable.

Table 5.2

Calculated and observed extinction distances for silicon

(Atomic Scattering Factors for electron waves taken from International Crystallographic Tables Vol. 3 Kynoch Press 1962)

Reflection	111	220	400	311
Calc. t_o(A)	724	885	1310	1600
Obs. t_o (unetched Faults)	690	860	not det.	
Obs. t_o (etched faults)	280	not det.		

Further support for the above conclusions comes from our observations that stacking faults occur almost invariably in closed figures or terminate in dislocation bundles rather than at single dislocations. The bundles themselves could be small closed prismatic figures as predicted.

The detection of stair-rod dislocations at the ends of faults is also in agreement with the observations that in many cases these dislocations lie parallel to <110> (which is a geometrical criterion for a stair-rod). However there are also cases where dislocations do not lie parallel to <110> and these can only be a/6 <112> partials. In the majority of cases these are observed as an incomplete part of an overlapping bundle e.g., at D Fig. 5.3b. The change in contrast along the faults e.g., in Figs. 5.1b, 5.2b, 5.3b is typical of many overlaps. However the detailed interpretation of contrast from such cases is difficult e.g., there are two possible interpretations for contrast. (1) For multiple faults of the same sign the contrast may change where the number of faults changes. Here the contrast reverses dark-light or vice versa and if fringes are visible they change by ±1 in number at each overlap. This must be the explanation of the change in contrast at D in Fig. 5.3b where the terminating dislocation does not lie along <110>. (2) A change in contrast will also occur if there is a change in spacing between overlapping faults without a change in number. This could be the case when one fault, ending at a stair-rod, steps up or down and continues closer to, or further from, its nearest neighbor. It is difficult to decide which causes contrast changes along a <110> direction, e.g., at F in Fig. 5.3b. There is obviously need for further theoretical treatment of contrast from overlapping stacking faults.

It will also be noticed that on many of the micrographs (Figs. 5.3b, 5.6a at G) there is a change in foil thickness extending for several microns along <110> from some of the stacking faults. This is shown by the darker contrast and, more convincingly, from the increase in the number of fringes at B, Fig. 5.6a. Since the foils are prepared by polishing from the substrate surface upwards (see Fig. 5.13) it follows that the top surface of the specimens cannot be plane. This thickening effect must be due to preferred growth on one side of certain faulted layers and would explain the contrast along the diagonals of the etch figures in Fig. 5.3a. This effect was seen mostly in [100] crystals.

It is known that the surfaces of epitaxial layers are often not plane but show very pronounced "growth pyramids". The nature and growth mechanism of these pyramids are not yet understood, and it may be that pyramids are associated with the observed phenomenon of preferential growth at the faults.

There are two further observations from the electron microscopy results that are of interest. One is that the faults were often preferentially etched as shown in Figs. 5.2b, c. This immediately suggests that the faults may have a different chemical composition from the adjacent parts of the crystal. Impurity segregation of oxygen and doping elements such as B, P, As etc., (from the substrate or the gas phase) to stacking faults is a possible explanation

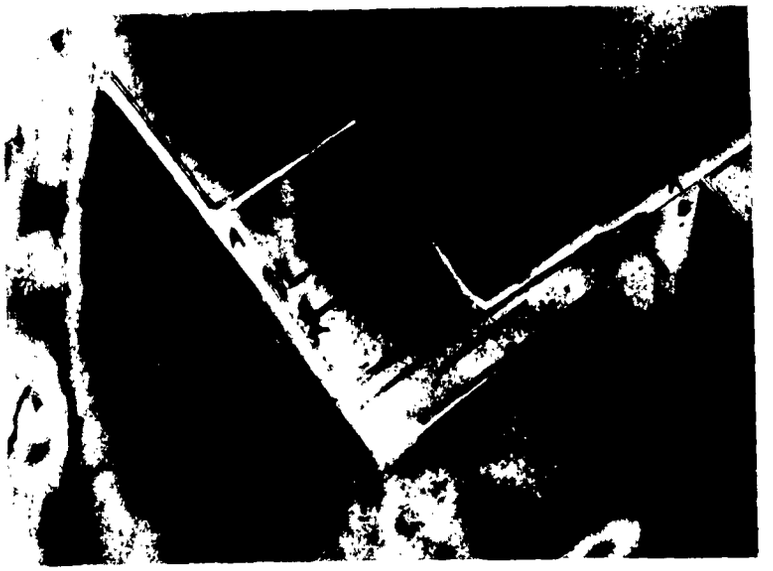


Figure 5.6a

Bright field image showing many overlapping faults bounding a twin A. Orientation [100].

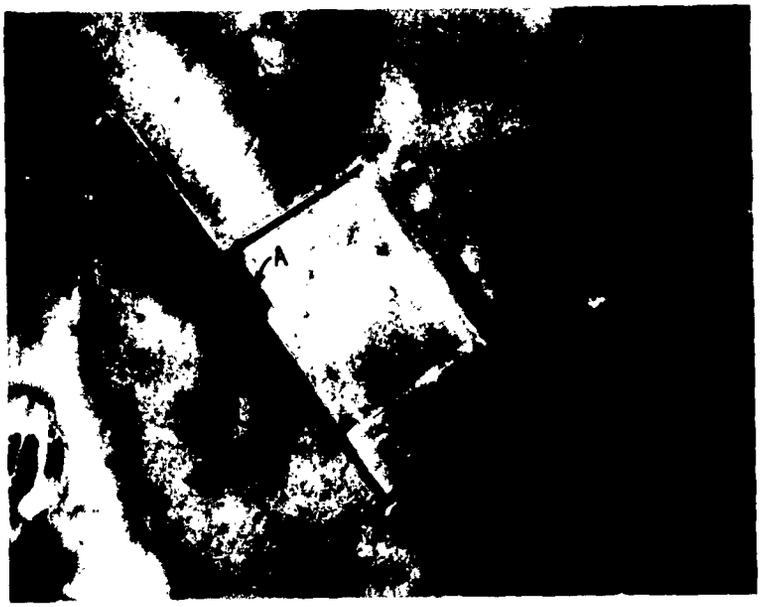


Figure 5.6b

Dark field image utilizing 400 reflection. Notice reversal of contrast at twins A.

of this effect. It is quite difficult to check this point, although we have some evidence in favor of impurity segregation from observations of the fringe spacings and comparison of the experimental and calculated values of the extinction distance t_o [65,66]. The results are shown in Table 5.2.

The low value of t_o for (111) in line 3 Table 5.2 can be accounted for by a change in composition at the faults as this would change the value of the scattering factor and, consequently, t_o since they are related by [66]:

$$t_o = \frac{\pi V \cos \theta}{\lambda F}$$

V volume unit cell

F structure factor

λ electron wavelength

Since t_o is very small for the etched figures F must be increased, e.g., due to the presence of elements of high atomic number. On this basis, As or P are more likely possibilities than O_2 or B. We should point out that accurate experimental determinations of t_o from the micrographs are difficult because many overlaps of faults may be present and this could affect the fringe spacings (e.g., depending on the number and separation between overlapping faults), and because we do not take into account any deviations of diffraction from the exact Bragg condition. However the fact that faults are clearly revealed in light microscopy after etching may also be taken as further evidence for impurity segregation.

The second observation was that of twinning, especially in the most heavily faulted crystals. Figure 5.6 shows examples of twins A in (a) bright field and (b) dark field images. The twin contrast in (b) can be further reversed by forming a dark field image of a twin reflection. The diffraction patterns and trace analyses show that these twins are in {111} confirming earlier work[39]. An interesting feature here is the lack of parallelism of the twin boundaries. This effect is probably due to many short steps along the boundaries, and must occur over small distances since it is not possible to resolve any steps.

(3) Origin and Growth of Faults

A unit area of stacking fault represents an excess of energy in the lattice. The magnitude of this stacking-fault-energy is important for creation and stability of the faults. From the frequent occurrence of twin boundaries and stacking faults in silicon one can conclude that the stacking fault energy in silicon should not be excessively high[67]. Phillips and Dash[68] have been able to create stacking faults by diffusing gold into single crystals of silicon. Stacking faults have also been observed in silicon precipitated out of a Si-Al melt[69]. No direct measurements of this energy have as yet been published in the literature.

---

67 A. Seeger, Handbuch d. Physik (Springer, Berlin 1955) vii, p. 639.

68 V. A. Phillips and W. C. Dash, J. Appl. Phys. 33, 568 (1962).

69 H. M. Bendler, J. Appl. Phys. 33, 240 (1962).

In the few, special cases where we have observed perfect dislocations and networks (see at A Fig. 5. 11b) no appreciable curvature at nodes could be detected. In electron microscopy, measurements of stacking fault energy are only possible when the radius of curvature of a node is measureably large[70]. In terms of stacking fault energy this would put the maximum value of γ which can be determined by this method at about 20 ergs/cm^2. In silicon therefore the stacking fault energy must be greater than this. Estimates[39, 71, 72] seem to indicate that the energy is of the order of 50 ergs/cm^2. This energy is comparatively low (Cu~40 ergs/cm^2, Al~180 ergs/cm^2). We conclude that energetically it is not improbable that growth stacking faults can be created during the epitaxial deposition. We shall now discuss the mechanisms of nucleation and growth of the stacking faults in epitaxial films.

It seems to be established that most, if not all, stacking faults originate at the interface between substrate and epitaxy films. This is proved by the equal size of the closed figures which increases with film thickness[61] and our observations from Fig. 5. 7 where the size AB of the many intersecting tetrahedra is always the same, to within about

---

70 A. Howie and P. R. Swann, Phil. Mag. 6, 1215 (1961).

71 R. Jaccodine, private communication (1962).

72 R. Siems, private communication (1962).

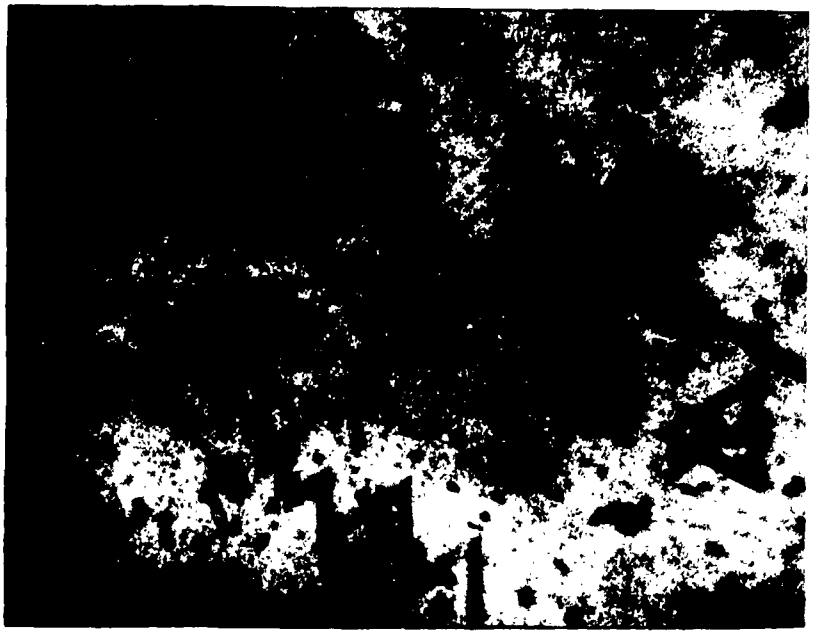


Figure 5.7

[111] foil showing intersection of at least 12 tetrahedra. The equality of the lengths AB prove that all the tetrahedra originate at the same depth i.e., at the substrate-film interface.

100 Å. In fact, it has been proposed [61] to use the dimensions of the triangles on (111) faces for a determination of the film thickness. No such conclusions can easily be reached for the dislocation-bounded fault ribbons, but it appears probable that these also nucleate preferentially at the interface: substrate-film. We shall now demonstrate that impurity contamination and mechanical damage of the substrate surface result in nucleation of stacking faults.

a. Impurity contamination

Previous authors offered the hypothesis that impurities at the interface between substrate and epitaxial deposit are important for nucleation of faults [50, 51, 60, 61, 62]. Evidence is presented here that oxygen contaminated substrate surfaces produce epitaxial layers with stacking faults.

Silicon samples that were well cleaned and heated for several minutes to temperatures above 1200 °C in hydrogen just prior to the epitaxial deposit did not yield layers with detectable stacking faults. This is evidence that the epitaxial growth procedure does not necessarily result in faulted material. However, when the specimens are boiled in water and not cleaned with HF or heat treated in hydrogen, vapor deposition gives consistently high concentrations of stacking faults (see section D).

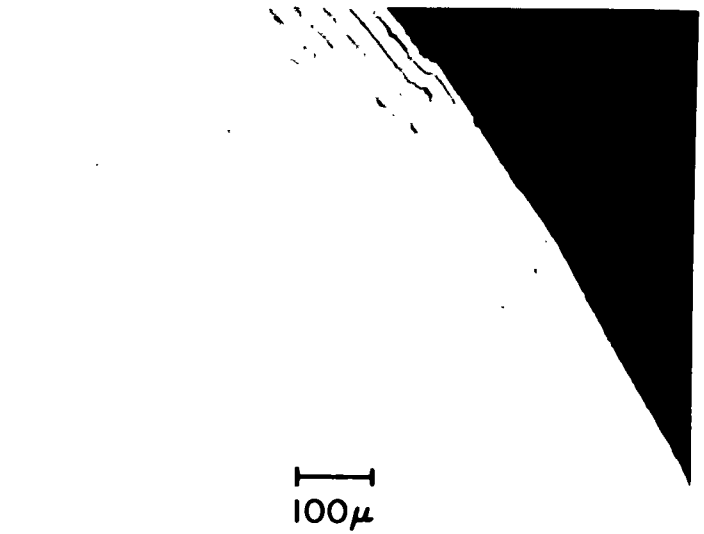
In order to perform a more quantitative study a different oxidation technique was used. Anodic oxidation is a valuable method to obtain oxide layers on silicon with good thickness control[73].

A series of slices with (100) surfaces were subjected to this anodic oxidation treatment for different times, resulting in different oxide thicknesses. The thickness of the oxide layer could not be determined, but they were estimated to be of the order of 100A or less. These slices were then used as substrates, together with an unoxidized, well cleaned control slice. The deposition onto these slices was made separately for each slice, starting with the control and proceeding in the order of increasing oxidation time. The slices were heated up quickly and the deposition was started as soon as possible in order to avoid complete removal of the oxide by reduction in the hydrogen stream.

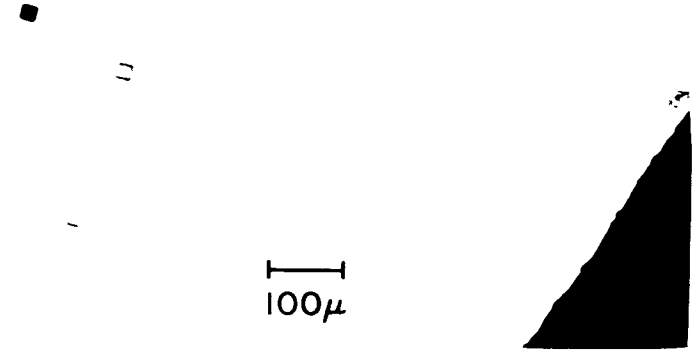
The results are reproduced in Fig. 5.8. A slice with a one minute oxidation gave the epitaxial layer depicted in Fig. 8a. A low density ($25/cm^2$) of stacking faults was observed, but it was definitely higher than the essentially fault-free control sample. Figure 5.8b shows increased fault density (approximately $55/cm^2$) for a heavier oxide layer. Still thicker oxide layers produced structures like the one shown in Fig. 5.8c. Very high fault densities were observed, but nonuniformly distributed over the slice. "Islands" of extremely

---

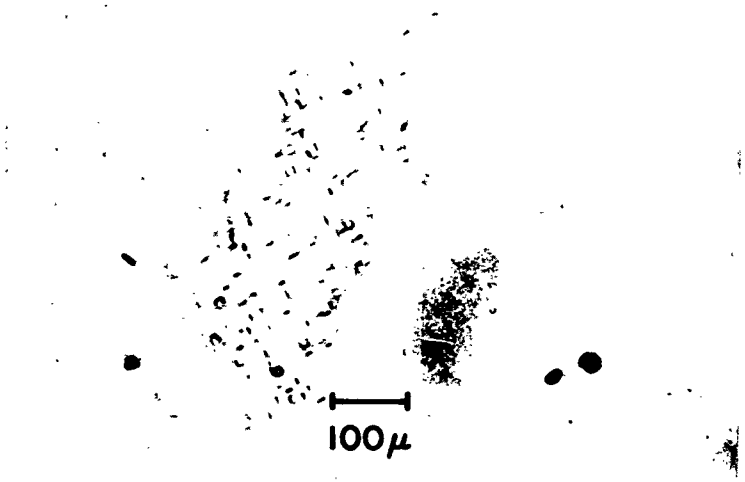
[73] P. F. Schmidt and W. Michel, J. Electrochem. Soc. 104, 230 (1957).



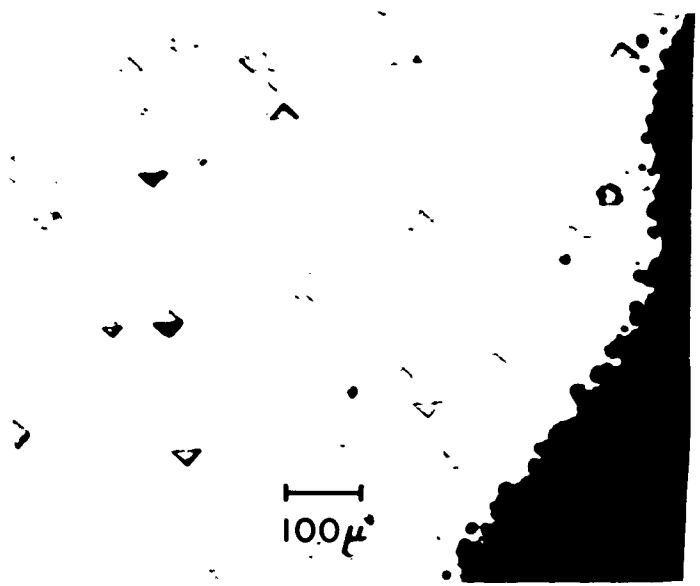
(a) Anodic oxidation for one minute



(b) 3 minutes



(c) 5 minutes



(d) 11 minutes

Figure 5.8

Influence of oxygen contamination on stacking fault nucleation in epitaxial silicon. (100) surface.

disturbed layers were found. This may be caused by nonuniform thickness of the oxide layer. Finally, for an oxidation time of 11 minutes, a severely damaged film resulted with a density of over $1000/cm^2$ of optically discernible faults. When it was attempted to deposit upon a thermally grown oxide layer with a thickness of the order of one micron, a polycrystalline growth was obtained (Fig. 5.9).

Corresponding electron micrographs prove that the increased density of defects shown in Fig. 5.8 paralleled an increased density of stacking faults.

b. Substrate damage

Mechanical damage of the substrate may play a role in stacking fault nucleation. This is demonstrated in an experiment with a scratched substrate surface. By scribing with a diamond stylus, loaded with 0.6 gms., the surface shown in Fig. 5.10a was obtained. After cleaning, the sample was used for epitaxial deposition. The etched surface is reproduced in Fig. 5.10b. High densities of faults were observed, particularly along the scribed lines. However, close inspection showed that stacking faults were also present over unscribed portions. Electron microscope observations showed that there were some regions having very high fault densities (Fig. 5.11a). Near these areas perfect dislocations and networks of random orientation

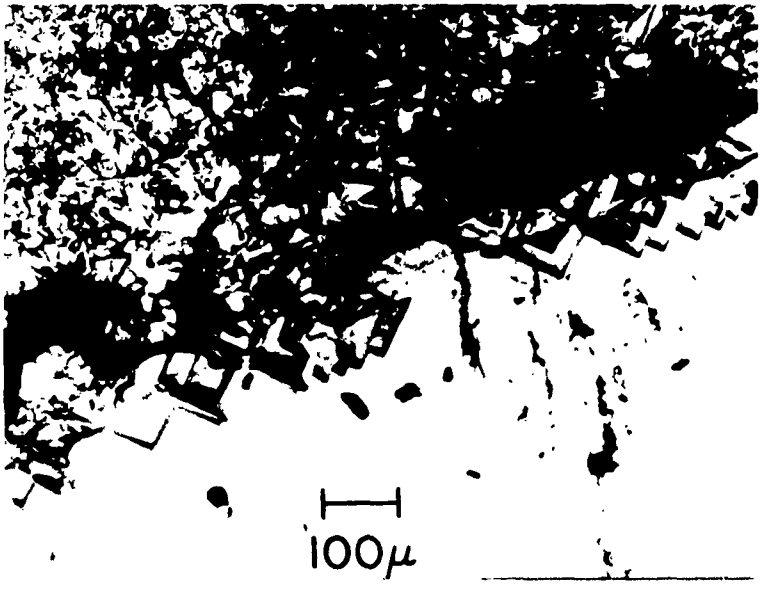
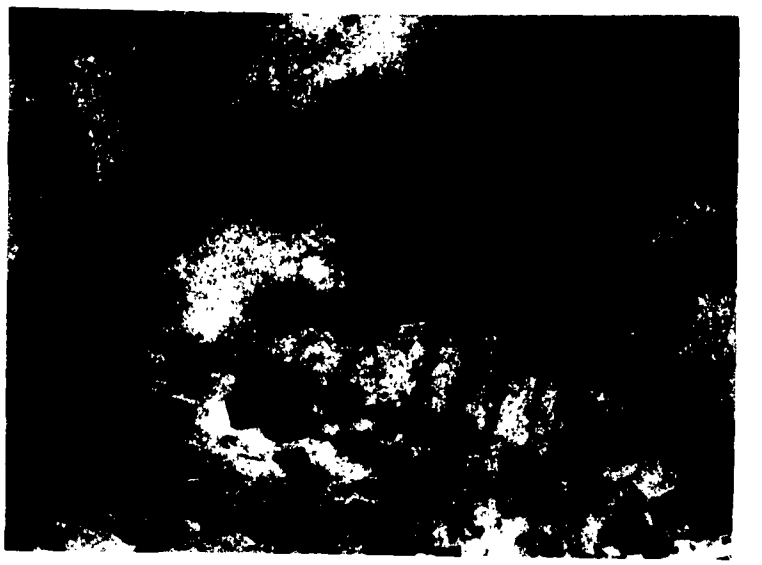


Figure 5.9

Epitaxial layer grown on a silicon slice half covered with a thermal oxide. The boundary is shown between that grown on the oxide, highly faulted, and that grown on the single crystal silicon.

(a)



(b)

Figure 5.10

Influence of mechanical damage on stacking fault nucleation, (100) surface.

were also observed (Fig. 5.11b). It is likely that these dislocations were inherited from the scribed substrate as they were never observed under any other conditions.

c. Influence of dislocations in substrate

Dislocations present in the substrate are known to propagate into epitaxial layers. This has, for example, been shown by the fact that edge dislocations in small angle grain boundaries propagate into epitaxial deposits[51,61]. Perfect dislocations, however, do not seem to constitute favorable nucleation points for stacking faults. It is possible to produce essentially stacking fault-free epitaxy material on silicon with dislocation densitites above 10^3 / cm^2. It is thus unlikely that a dislocation reaction is causing the creation of two partials plus one stacking fault.

D. Mechanism of Fault Formation

The experiments showed that incomplete removal of oxide from the surface of the substrate crystal prior to the start of epitaxial growth was the most important single cause of stacking faults in the deposited layer. Patches of oxide can produce steps at the surface of the substrate crystal the height of which may not be equal to an integral multiple of the appropriate interplanar spacing. When the silicon crystal is forced to grow around such an oxide inclusion, coherence may be maintained by the formation of stacking faults. Figure 5.12 is a (110) section through a model drawn to scale of a layer that has been deposited on a

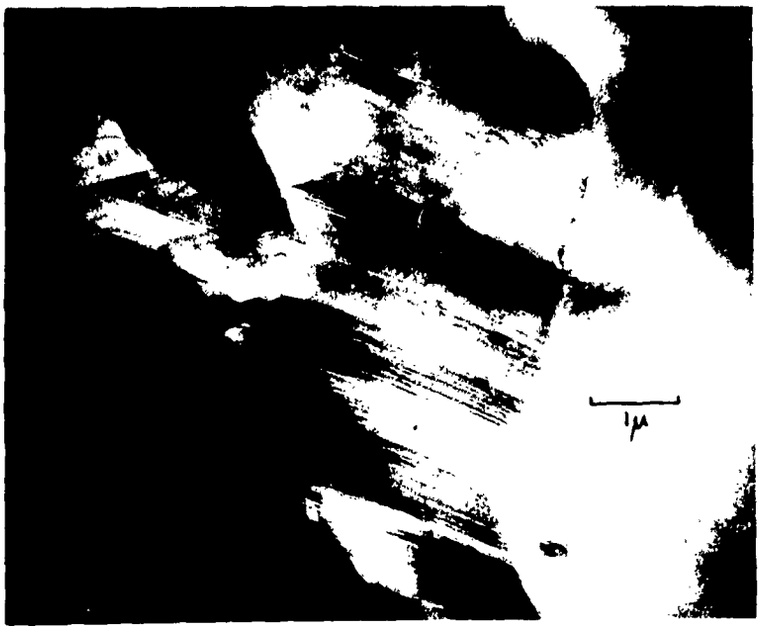


Figure 5.11a

Showing very high density of faults and twins in foil prepared from sample shown in 5.9b. [100] foil.

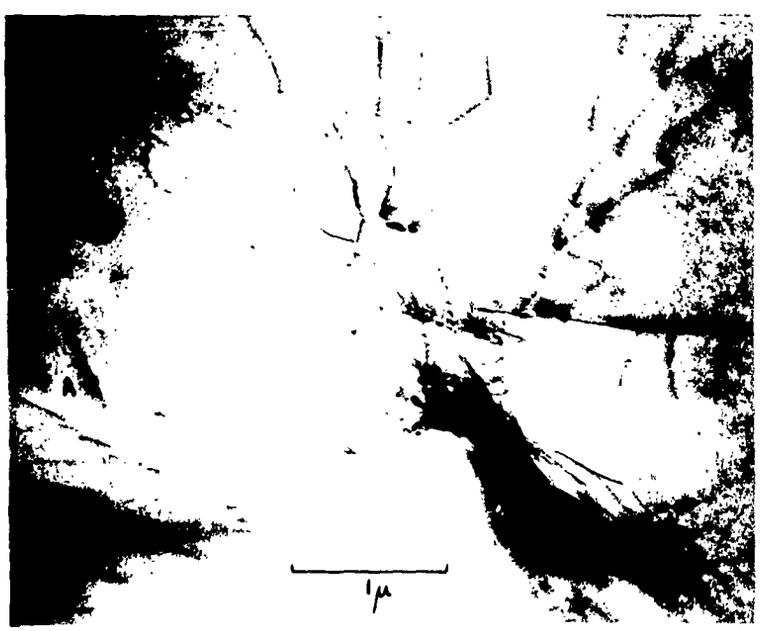


Figure 5.11b

As (a) showing undissociated dislocations lying in traces of 111 .

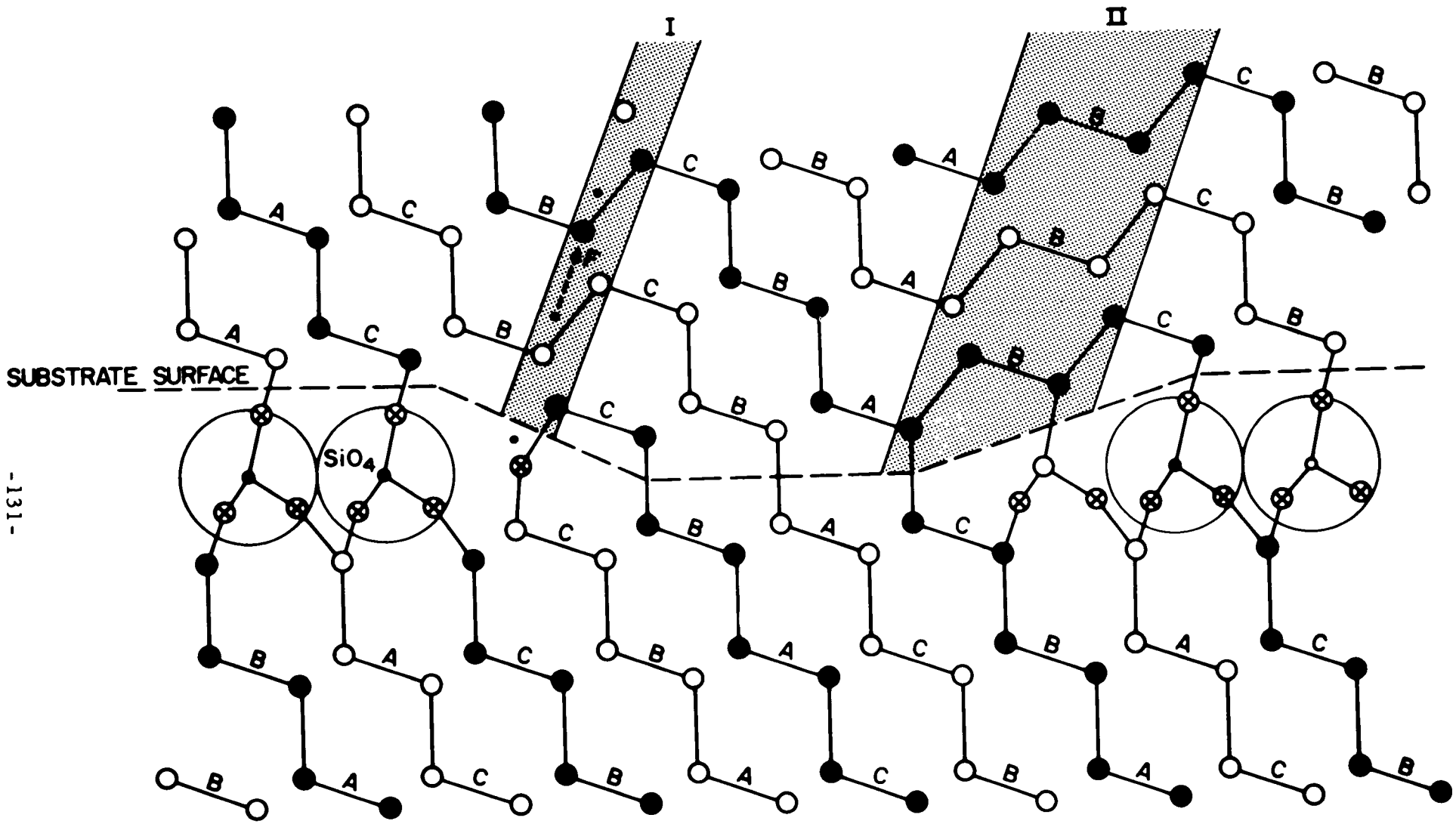


Figure 5.12

(110) section through a model drawn to scale of an epitaxial layer that has been deposited on a partially oxidized (111) surface showing accomodation at edges of oxide layers by introduction of stacking faults.

(111) surface. The original substrate surface, partly covered by a thin oxide layer, is indicated by the dashed line. In the model it has been assumed that the oxide layer corresponded to replacement of a single layer of silicon atoms by SiO_4 tetrahedra distorted in such a way as to maintain perfect coherence with the silicon. The Si-O-Si and Si-Si distances were taken as 3.05 Å and 2.35 Å respectively[74]. With these assumptions it can be seen that coherence around the oxide patch is almost perfectly maintained by introducing stacking faults that start at the edges of the oxide and grow into the deposited layers of silicon along the {111} planes. For a given (111) plane the fault must be either the intrinsic or extrinsic type depending on the sense of the step as shown at I and II respectively. The shear necessary for the intrinsic fault is indicated by the arrow F in Fig. 5.12.

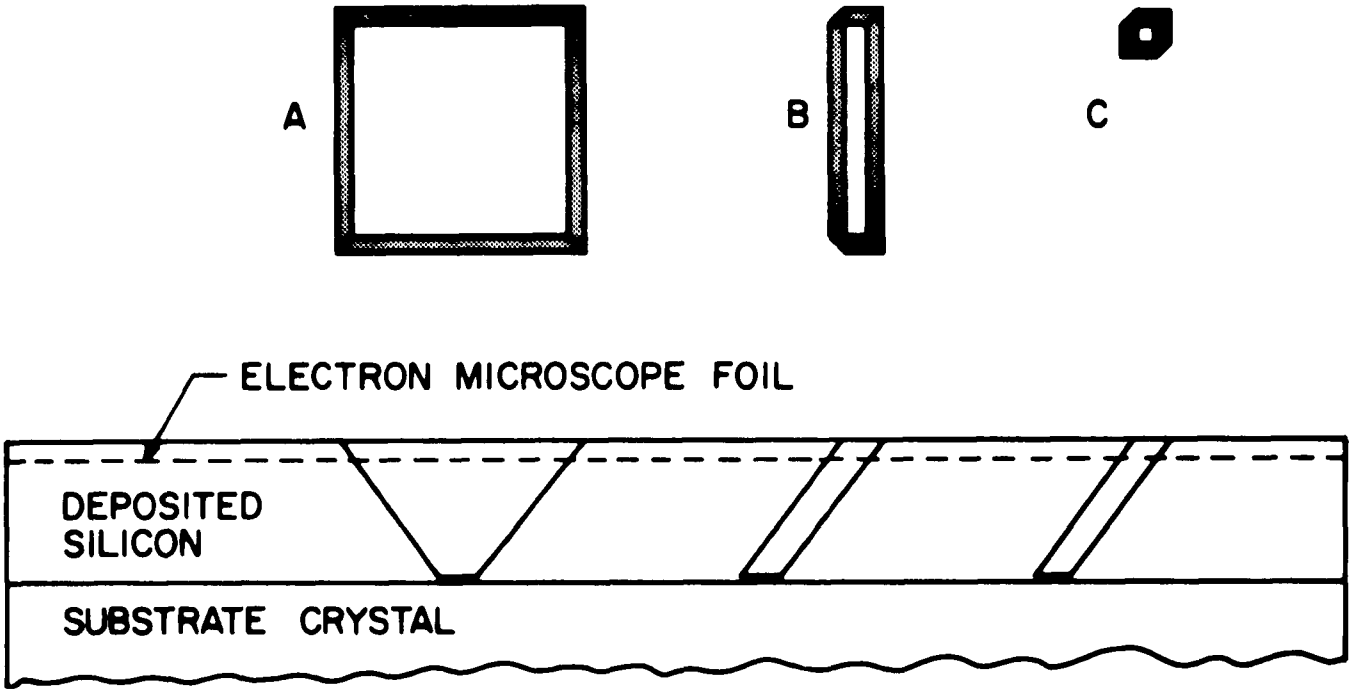
If both types of faults are formed, several different configurations of faults could originate at a small patch of oxide or at a small hole in an almost continuous oxide layer. Faults on three {111} planes would be nucleated for (111) or (110) surface orientations and on four {111} planes for a (100) substrate. Three different configurations as they would appear in a thin foil taken near the top surface of a thick deposited layer of (100) orientation are shown in Fig. 5.13a. Square figures like that at A were very often observed in the electron micrographs. This configuration

---

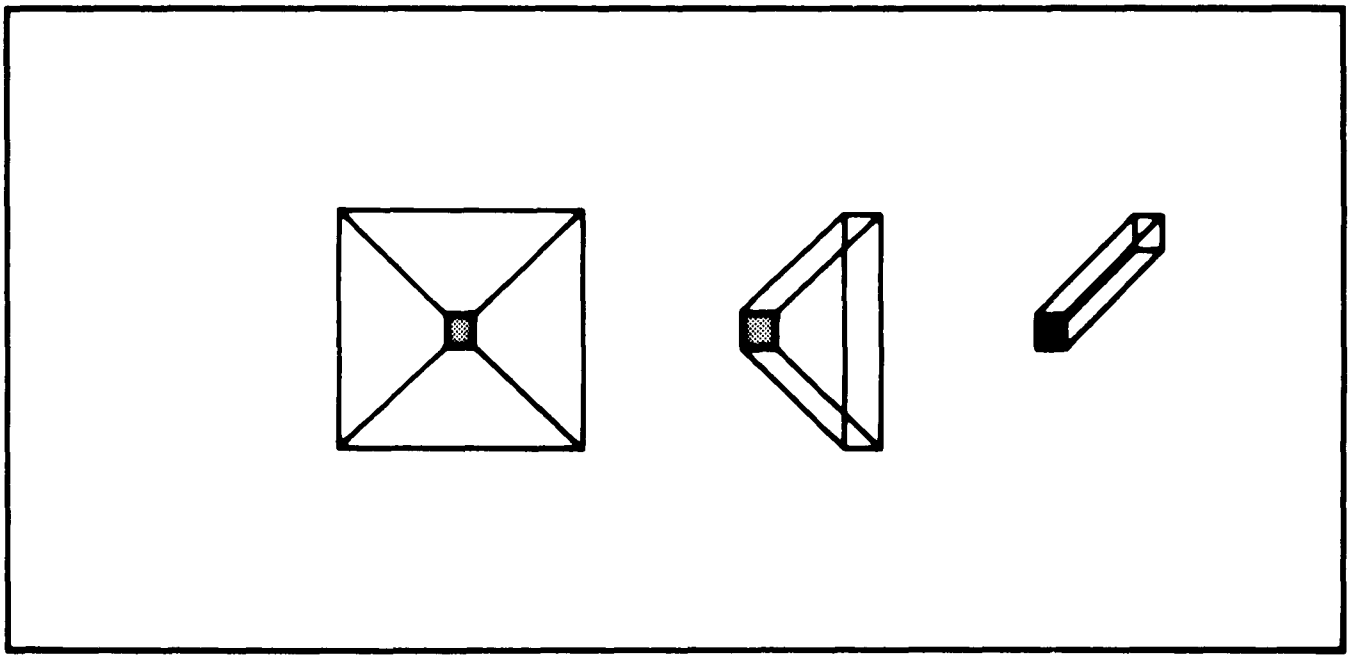
[74] R. C. Evans, Introduction to Crystal Chemistry, Cambridge University Press, 1948, p. 242.

results when all the four faults are of the same type. If only three faults are of the same type, a long rectangular figure as shown at B is formed. The small square figure at C corresponds to two faults of each type. Electron microscope foils were always thin compared to the total thickness of the deposited silicon as indicated by the dashed line in Fig. 5. 13b. The top and side views of the deposited material show how the three configurations could originate at similar patches of oxide at the substrate surface. In Fig. 5. 13 the size of the oxide patch is assumed to be larger than the thickness of the electron microscope foil.

If the size of the oxide patch was small (less than a few hundred angstroms) compared to the thickness of the electron microscope foil there would be no noticeable change in the configuration, A , but the extrinsic-intrinsic pairs in B and C would then be close enough to be unresolvable. The configuration shown at B would appear to be a single fault with a length equal to one side of the square A as at X in Fig. 5. 1a. The small figure, C , would appear only as a line of the same length and direction as one of the stair-rod dislocations at the corners in A. Actually it would still be a small bundle of four parallel stair-rod dislocations connected by four short stacking faults. Possible examples are at E, Fig. 5. 3b. The electron diffraction contrast effects described previously suggest that some of the apparently single faults may actually be intrinsic-extrinsic pairs.



(a) Faults as they would appear in electron microscope foil



(b) Section through substrate and deposited layer showing relation of stacking fault figures to nucleating oxide inclusion.

Figure 5.13

Three stacking fault configurations that could be nucleated by a small patch of oxide or a small hole in an almost complete oxide layer.

Intersection of a number of pyramid stacking fault configurations as the thickness of the foil increases can produce complex zig-zag fault systems. In cases where it was possible to make accurate measurements of the dimensions of a large number of close by closed figures, (Fig. 5.7), it was found that they were identical to within 100Å. This shows that they must have all originated exactly at the original substrate surface and that the nuclei were either all of the same size or all less than 100Å in diameter.

It was also found that closed stacking fault figures were not always single faults but sometimes were unresolved multiple faults of the same type (either two or more parallel intrinsic faults or more than one extrinsic fault). For example this must be true for the triangular figure in Fig. 5.1b. Multiple faults might be nucleated if there were steps in the thickness of an oxide inclusion; coherence being maintained at each step by introduction of another fault.

E. Conclusions

Epitaxial methods are important for the production of silicon devices. This work has shown that under certain conditions, high densities of stacking faults can occur in epitaxially grown layers. Analogous to the weak direct effects of dislocations upon electrical device parameters, one also expects some disturbances due to stacking faults. P-n junctions in semiconductors containing stacking faults fail to exhibit anomalies which can be directly attributed to faulting. However, segregation of impurities may lead to appre-

ciable effects. Undesirable "soft" reverse characteristics of p-n junctions can be caused by metallic precipitates[14], which might possibly also change the forward junction-characteristic[47].

Immediately prior to the termination of this contract, evidence was found for an electrical effect at stacking faults[75]. Microplasma breakdown of reverse biased p-n junctions in epitaxial silicon with stacking faults was studied*.

Figure 5.14 shows a photograph of a guard-ring n^+p^- diode in epitaxial silicon. The triangular figures are stacking faults which were deliberately introduced by oxide contamination of the substrate surface. The light spots shown are microplasmas which are observed when the diode is reverse biased. The corners of some stacking fault figures are seen to constitute sites of preferential breakdown. It is logical to correlate the microplasmas with the stair-rod dislocations. It has been shown before that several bends, i.e. several stair-rods, are often present at the triangle corners, which causes more than one microplasma. This is in agreement with the electron micrographs, as discussed earlier. The fact that not all corners of the faults show microplasmas seems to be a confirmation of the hypothesis that the

---

75 H. J. Queisser and A. Goetzberger, Bull. Am. Phys. Soc. 7, 602 (1962) Stanford Meeting APS, December 1962.

* Work done in cooperation with A. Goetzberger and others, partly sponsored by Air Force Cambridge Laboratories, Contract AF 19 (604) 8060.

Figure 5.14

Electrical Disturbances Associated with
Stacking Faults in Epitaxial Silicon

dislocation does not cause microplasma breakdown directly, but rather indirectly, such as by precipitates.

This result shows that the stacking faults are indeed important structural defects with serious consequences for the performance and reliability of p-n junction devices which consist of epitaxial semiconductor material.

Part II

THERMAL INSTABILITY IN EXTENDED TRANSISTOR AND THERMISTOR STRUCTURES

1. INTRODUCTION

This part of the report deals with the problem of lateral thermal instability in transistors. It presents the basic theory on how thermal instability arises in thermistor and transistor structures and shows experimental evidence for instability with consequent development of "hot spots" in silicon power transistors.

The instability referred to here is a lateral instability in which current builds up in one region of the device and decreases in the remaining area so that a region of high local current density and high temperature is produced. This can occur even if the external current is kept constant. Under the right circumstances, the localized high temperature region can generate enough carriers thermally to provide sufficient internal base current to maintain the collector current in the hot spot. Under these conditions no external base current is required, and the transistor remains turned on if the base is opened. It is believed that this situation produces the phenomenon known as "secondary breakdown."[1] Secondary breakdown is recognized as being a serious reliability

---

[1] H. A. Schafft and J. C. French, IRE Transactions on Electron Devices, Vol. ED-9, p. 129 (1962). (References)

problem in power transistors and to avoid its effects transistors must generally be operated far below their potential power dissipation capabilities.

To get a simple physical picture of lateral thermal instability, consider a transistor with a uniform emitter-base voltage and constant external emitter current. A reverse bias is applied to the collector junction so that the transistor is dissipating power and in consequence the active region of the transistor rises to a higher temperature than the heat sink. The temperature rise depends on the power being dissipated and the thermal conductance of the transistor. Now suppose that the current density is not uniform but is larger in some localized part of the structure. This increase, for example, may be produced by a defect of some sort such as a thin spot in the base layer. A local temperature rise will then occur. At constant forward bias across the emitter junction the temperature coefficient of current across the emitter junction for a silicon transistor at representative operating bias is such that about a 10% increase in current occurs for every degree centigrade rise in temperature. Thus, the local temperature rise will produce a further increase in local current and one has a positive-feed-back-like situation. If the "gain" around this feedback loop is greater than one, that is if the temperature rise produces an increase in current larger than the initial current increment originally causing the temperature rise, an unstable situation exists. Current and temperature will build up locally until they are limited by resistive drops in the various layers of the transistor or by changes in state of the materials such as melting, with resulting device destruction.

We shall first develop a theoretical treatment of lateral thermal instability. The simple case of a thermistor is considered first, two parallel thermistors next. A treatment of the distributed thermistor or transistor structure follows. Current and temperature distributions in interdigitated structures are discussed subsequently. This is considered important because the experimental evidence was obtained on an interdigitated silicon power transistor structure. The last chapter of the section on theory deals with the localization of "hot spots." Such localization is thought to arise mainly from structural defects, such as the ones discussed in the first part of the report.

An account of the experimental results is given after the theoretical analysis. Electrical and thermal measurements are cited as evidence for the formation of localized "hot spots." The development of these hot spots, the factors influencing their formation, and the resulting failure and damage are described.

2. THE STABILITY INDEX s FOR A THERMISTOR

In order to introduce the influence of heating upon thermal instability we consider first the case of a thermistor in which the resistance all arises from a small region taken to be at a single temperature. We shall consider small deviations in current, voltage and temperature from a set of initial values. The symbols used are as follows:

$$V_1 = V_0 + v \tag{2.1a}$$

$$I_1 = I_0 + i \tag{2.1b}$$

$$U_1 = U_0 + u \tag{2.1c}$$

where V_0, I_0, and U_0 represent reference (usually steady state) values of voltage current and temperature, with temperature defined in equivalent voltage as

$$U = kT/q \tag{2.2}$$

where k is Boltzmann's constant, T temperature on the Kelvin scale, and q the electronic charge. The small letter symbols represent disturbances in the linear range of the phenomena investigated, and the quantities with subscript 1 represent the changed values.

Two functional relationships exist among the quantities considered:

$$I_1 = I_0 + i = F(V_1, U_1) = I_0 + gv + I_0 au \qquad (2.3)$$

$$dE_T(U_1)/dt = K\dot{u} = V_1 I_1 \quad H(U_1)$$

$$= V_0 i + I_0 v - hu \qquad (2.4)$$

The first equation states that the current is a function of voltage and temperature; in the linear range of small disturbances its change i is represented by a conductance g and a temperature coefficient a, so that

$$i = gv + I_0 au \qquad (2.5)$$

The second equation is the law of conservation of energy, which states that the rate of storage of heat energy E_T (which is the heat capacity K times rate of temperature rise $\dot{u}$) is equal to the net power $V_1 I_1$ minus $H(U_1)$ the rate of heat conduction to a heat sink at temperature $T_s = qU_s/k$. At the steady state condition, these powers $H(U_0)$ and $V_0 I_0$ exactly balance so that Eq. (2.6) involves only disturbances from equilibrium:

$$K\dot{u} = V_0 i + I_0 v - hu \qquad (2.6)$$

where in the linear range

$$hu = H(U_1) - H(U_0) = udH/dU \qquad (2.7)$$

We shall develop below an impedance formula based on (2.5) and (2.6). As will be shown below, the existence of a negative real part for the impedance depends upon a value greater than unity for a stability

index s, defined as

$$s \equiv a V_0 I_0 / h = (U_0 - U_s) a \qquad (2.8)$$

To a first approximation $V_0 I_0 / h$ is the rise in temperature in U-type units above the heat sink and is accurately this rise if heat flow is linear in temperature gradient so that

$$H(U) = (U - U_s) h \qquad (2.9)$$

If the current depends exponentially upon temperature rises, then s = 1 corresponds to a current increase at fixed voltage by a factor of e = exp(1), the base of the natural logarithms.

It should be noted that a/h is a quantity independent of the temperature scale. It has the dimensions of Watt^{-1}. Thus, although the U in volts is algebraically more convenient to use in equations (partly because t is also used for time) than is T in °K, the quantity a/h may equally well be calculated in terms of a_T, the temperature coefficient in $(°K)^{-1}$ of current increase at constant voltage, and of h_T, heat conductance in watts per °K. The relationship is

$$a_T / h_T = (a k/q) / (h k/q) = a/h \qquad (2.10)$$

It is frequently more convenient to calculate s in terms of the a_T and h_T units:

$$s = a_T V_0 I_0 / h_T = (T_0 - T_s) a_T \qquad (2.11)$$

where T_0 and T_s correspond to U_0 and U_s.

In order to understand why the stability index s is a critically important quantity, we consider first the case of zero frequency. Fig. 2.1 shows a V-I plot of equations (2.4) and (2.5) for two temperatures U_0 and U_1. The steady state condition for temperature U_0 is represented by the V_0, I_0 pair that simultaneously satisfies the two equations. Similarly, V_1 and I_1 satisfy the two equations for the higher temperature $U_1 = U_0 + u$. For the case shown, both v and i are positive and the thermistor exhibits a positive low frequency conductance.

The particular example chosen corresponds to the important situation in which at a constant temperature a positive differential conductance arises from I = F(V, U) of Eq. (2.3); this is represented on Fig. 2.1 by the pair of curves I = F(V, U); these also represent the case of a positive temperature coefficient of current at constant voltage. In fact, at constant voltage the current increases from I_0 to $I_0 + I_0 au$ and the increase in power input will be $V_0 I_0 a u$ corresponding to the point at the end of the dashed arrow. The significance of the fact that this arrow does not reach $V_1 I_1$ is discussed below.

The power inputs required for the two temperatures U_0 and U_1 are represented by two hyperbolae, each with constant input power VI, as shown in Fig. 2. For the case represented, it is seen that the point at the end of the dashed arrow at constant $V = V_1$ does not lead to the necessary power to produce a temperature rise u. It is evident that to reach $V_1 I_1$ an

increase in V as well as in I is necessary.

If, however, the $H(U_1)$ curve were much closer to the $H(U_0)$ curve (a situation which would arise for high thermal resistance so that only a small increase in power produces a large temperature rise), then the arrow at constant V would lead to too much power and the V_1I_1 point would correspond to an increase in I and a decrease in V, and a negative differential steady-state conductance.

That the dividing condition between positive and negative conductance corresponds to $s = 1$ may be seen by noting that in order for the increase in power due to the current increase $I_0 a u$ to be sufficient to sustain the increased temperature without any change in V, the relationship

$$V_0 I_0 a u = h u \tag{2.12}$$

must hold so that

$$s = a V_0 I_0 / h = 1 \tag{2.13}$$

This condition that $s = 1$ is a simple example of "self-balance" discussed in Section 5; when self-balance occurs, a disturbance in temperature produces at constant voltage exactly the change in power needed to compensate for the change in heat flow.

Two limiting cases for the thermal conductance h are of interest. These are represented in Fig. 2.2. In one case it is assumed that $s \gg 1$ so that the increase i at $v = 0$ for an increase in u is much larger than necessary to give the power required to produce u. This corresponds to

a relatively low heat conductance. As the differential heat conductance h approaches zero, the H-family of curves approach a single hyperbola of constant power and large temperature rises, represented by U_2 and U_3, are produced by relatively small increases in power. Consequently, the V-I characteristic approaches the constant power hyperbola. (A practical test to see if an observed negative resistance is due to a high temperature coefficient resistance and poor heat sinking is to see if $dV/dI \doteq - V/I$, the consequence of $VI \doteq$ const.)

The other extreme limit corresponds to $s \ll 1$. In this case relatively large changes in power produce relatively small increases in temperature represented by U_1 and U_2 and as h approaches infinity and s approaches zero, the V-I characteristic approaches the isothermal $I = F(V, U_0)$ curve.

An algebraic formula for differential impedance at circular frequency ω may be readily obtained for the model under consideration by eliminating u from equations (2.3) and (2.6), which we rewrite as follows:

$$i = gv + I_0 a u \tag{2.14}$$

$$j\omega K u = V_0 i + I_0 v - h u \tag{2.15}$$

The result may be expressed in the form

$$\begin{aligned} v/i &= (1 - s + j\omega\tau)/(1 + s' + j\omega\tau)g \\ &= r - r'/(1 + j\omega\tau') \\ &= [g - g''/(1 + j\omega\tau'')]^{-1} \end{aligned} \tag{2.16}$$

where the symbols have the following meaning:

$$s = a V_0 I_0 / h \tag{2.17a}$$

$$s' = s I_0 / V_0 g \tag{2.17b}$$

$$r = 1/g \tag{2.17c}$$

$$\tau = K / h \tag{2.17d}$$

$$\tau' = \tau / (1 + s') \tag{2.17e}$$

$$r' = r(s + s') / (1 + s') \tag{2.17f}$$

$$\tau'' = \tau / (1 - s) \tag{2.17g}$$

$$g'' = g(s + s') / (s - 1) \tag{2.17h}$$

The impedance expressions of Eq. (2.16) correspond to various possible equivalent circuits of which Fig. 2.3 presents two examples. These may be helpful in visualizing for what terminal conditions the circuit will be unstable.

One particularly important case occurs when the equivalent circuit is connected to a zero impedance external circuit. Under these conditions it is seen that the negative capacitor is shunted by two conductances in parallel. If the positive conductance $1/r$ is larger in magnitude than the negative conductance $-1/r'$, an exponential build-up will occur for the voltage across the condenser and the current i through the external circuit. The condition for this exponential build-up is that r is less than r', which is identical with the condition that the stability index s be greater than unity. We shall return to this stability condition in connection with two thermistors in parallel.

For large temperature rises there may be failure of the linear approximations leading to s = 1 for the stability condition. A more general condition can readily be derived from the self-balance condition. If it is assumed that a fluctuation produces an increase u above the steady state value U_0, then if the additional heat generated electrically at constant voltage exceeds the additional heat flow due to increased temperature, the fluctuation will grow and instability will occur. Self-balance is the dividing condition between stability and instability and can be expressed conveniently in terms of fractional changes in electrical and thermal powers, since the steady state powers I_0V_0 and $H(U_0)$ are equal. Fractional changes are, of course, equal to changes in logarithms. Furthermore, since V_0 is constant the change in log V_0I_0 equals the change in I_0. Hence, taking rates of change instead of changes themselves, we obtain

$$\partial \log I / \partial \log U \leq \partial \log H(U) / \partial \log U \qquad (2.18)$$

where the derivatives are evaluated at U_0, I_0 and V_0. When $H(U_0)$ is proportional to the rise above the heat sink temperature U_s, this condition reduces to

$$a U_0 \leq U_0 / (U_0 - U_s) \qquad (2.19)$$

or

$$(U_0 - U_s)a \leq 1 \qquad (2.20)$$

in keeping with equations (2.8) and (2.9).

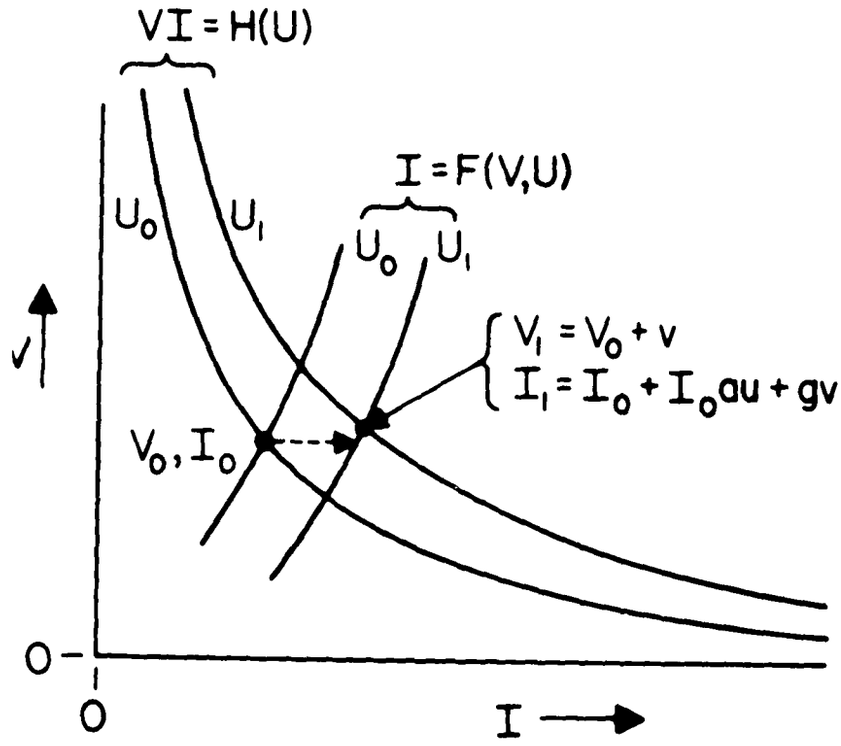
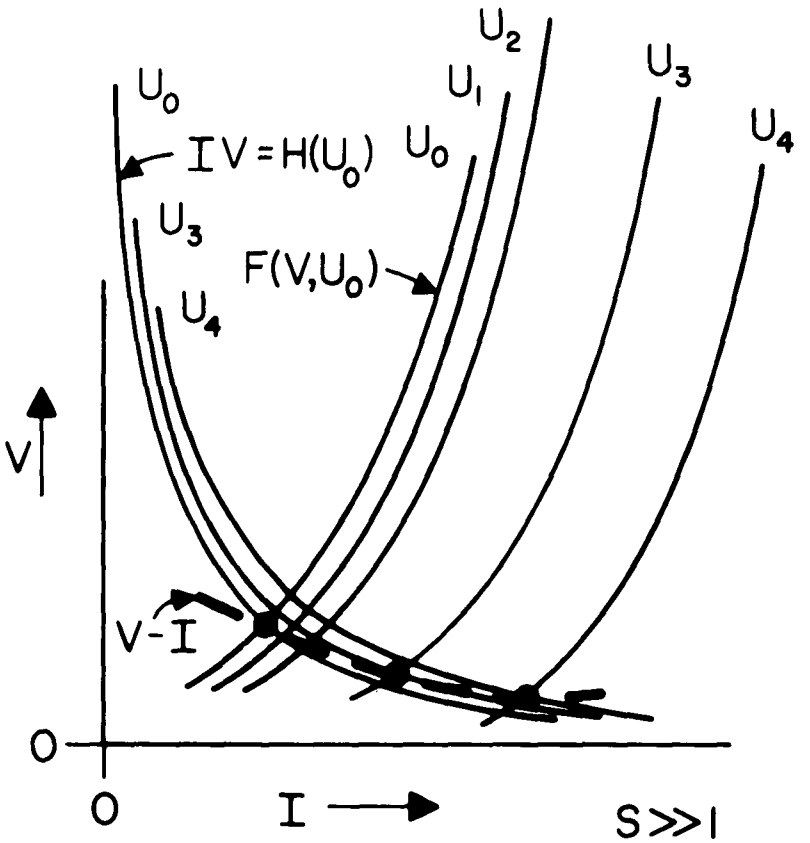


Figure 2.1

Relationship of two steady state conditions.



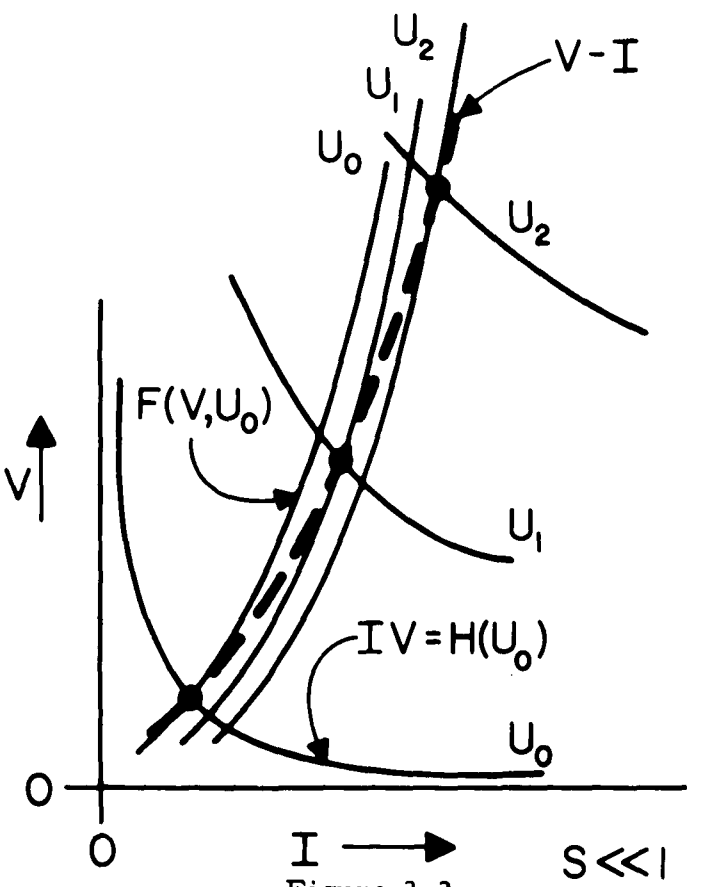


Figure 2.2

Illustrative limiting cases of low and high thermal conductance

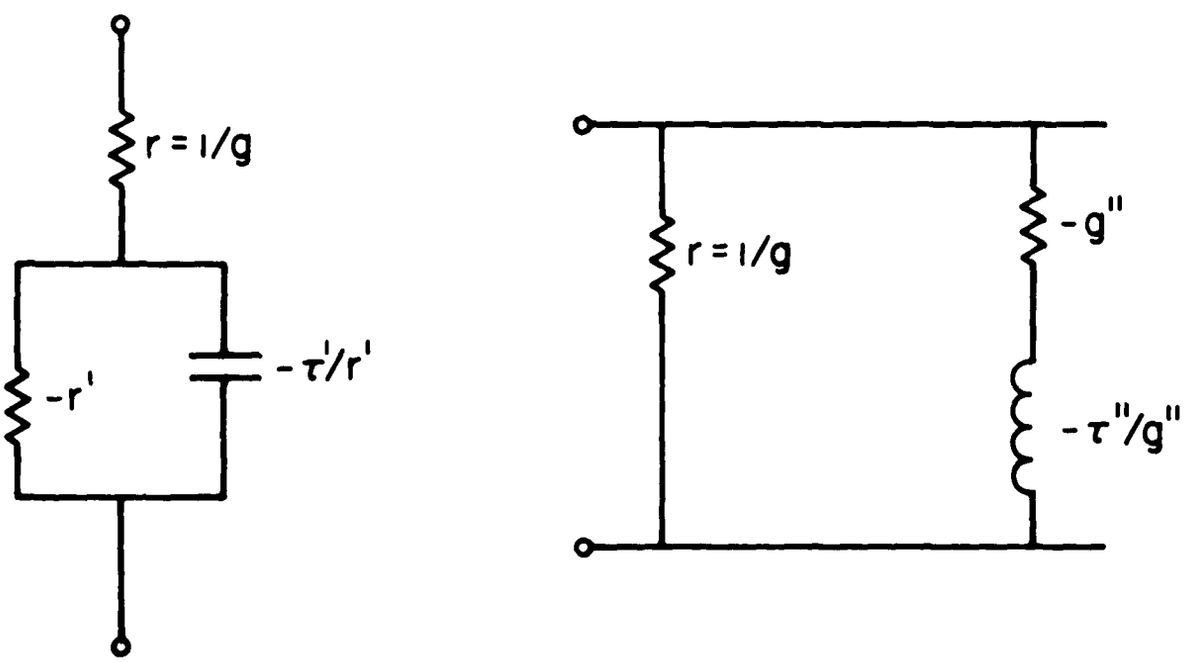


Figure 2.3

Small signal equivalent circuits for thermistor.

3. THE TRANSISTOR AS A THERMISTOR

This study is chiefly concerned with thermal instability caused by the increases in current through the base layer of a junction transistor produced by increase of temperature. This effect causes the transistor to simulate a thermistor having a high positive temperature coefficient of conductance.

Fig. 3.1 illustrates the configuration studied in this section. An npn transistor is represented. It is represented essentially as a two-terminal device carrying a current I with an applied voltage V. Internal to the device there is a base voltage supply V_b which maintains the base voltage. For most cases of practical interest the power dissipation due to the base current and the base emitter voltage is negligible. In special cases this power dissipation may have to be considered, but we shall in general neglect it here.

We shall concentrate our attention on the circumstances in which the collector current is saturated so that it is substantially independent of collector voltage. This means that the family of I = F(V, U) curves discussed in Section 2 are nearly parallel to the voltage axis. As is seen in Fig. 3.1, the constructions of Fig. 2.2 can lead to negative resistances for the V-I characteristic.

In analyzing the behavior of the transistor in the circuit of Fig. 3.1, we shall consider how the collector current depends upon temperature and base voltage and also how the base voltage depends upon temperature for constant collector current. We shall neglect recombination in the base layer in this treatment, so that the current of importance is the collector current, denoted by I. As a starting point we shall assume that a uniform forward bias V_b exists over the area A of the emitter-base junction. If the total charge of uncompensated acceptors per square centimeter of the base layer is Q, and the diffusion constant for electrons is D_n, then this current may be expressed as

$$I = (A q^2 n_i^2 D_n / Q) \exp (V_b / U) \qquad (3.1)$$

where q is the electronic charge and n_i is the electron or hole density in intrinsic material. U is the temperature expressed in electron volts:

$$U = kT / q \quad ; \quad T/U = q/k = 11,600°K/volt \qquad (3.2)$$

where k is Boltzmann's constant. In this expression n_i and D_n are both dependent upon temperature. Over a range of temperature sufficiently broad to cover the studies undertaken here, the dependences of n_i^2 and D_n upon temperature can be expressed as follows:

$$\ln D_n = \ln (kT\mu_n / q) = (1 + n_{\mu n}) \ln T + \text{const} \qquad (3.3)$$

$$\ln n_i^2 = n_{pn} \ln T - (V_g / U) + \text{const} \qquad (3.4)$$

where the mobility of electrons varies as $T^{-n_{\mu n}}$.

Combining these expressions with expression (3.1) leads to

$$\ell n\, I = n_T\, \ell n\, U + (V_b - V_g)/U + \text{const} \tag{3.5}$$

where the power n_T is given by

$$n_T = 1 + n_{\mu n} + n_{pn} \tag{3.6}$$

The desired derivatives of current in respect to voltage and temperature are readily obtained from Eq. (3.5). It is mathematically convenient to express these in terms of logarithmic derivatives which can be readily altered to give derivatives in respect to the variables involved. The total differential of Eq. (3.5) is

$$d\,\ell n\, I = [n_T + (V_g - V_b)/U]\; d\,\ell n\, U + d\,V_b/U \tag{3.7}$$

The partial derivative of $\ell n\, I$ in respect to $\ell n\, U$ at constant V_b is readily found to be

$$a_o \equiv \left(\frac{\partial\, \ell n\, I}{\partial\, \ell n\, U}\right)_{V_b} = (V_g - V_b)/U + n_T \tag{3.8}$$

This quantity, which has no dimensions, would be the slope of the current versus temperature line on a sheet of log-log graph paper. It is related to the coefficient a, discussed in the previous section, as follows.

$$a \equiv \left(\frac{\partial\, \ell n\, I}{\partial\, U}\right)_{V_b} = \left(\frac{\partial\, \ell n\, I}{\partial\, \ell n\, U}\right)_{V_b} \div U \equiv a_o/U \tag{3.9}$$

Since $\ell n\, U$ and $\ell n\, T$ differ by a constant, it may readily be seen that the coefficient of current change in respect to temperature in degrees

Kelvin is given by

$$a_T = \left(\frac{\partial \ell n I}{\partial T}\right)_{V_b} = a_o / T \tag{3.10}$$

Also of interest is the derivative of current in respect to base voltage at constant temperature, and this is found from Eq. (3.7) to have the familiar form

$$\left(\frac{\partial \ell n I}{\partial V_b}\right)_U = 1/U = q/kT \tag{3.11}$$

One remaining derivative of interest in the three variable formula of Eq. (3.1) is the derivative of base voltage in respect to temperature at constant current. This derivative is used conveniently in estimating temperature rises in junction transistors by finding the change in the emitter-base voltage when the current is held constant. This derivative is obtained from Eq. (3.7) by taking $d\,\ell n\, I = 0$, the result being

$$\left(\frac{\partial V_b}{\partial T}\right)_I = \left(\frac{\partial V_b}{\partial U}\right)_I \frac{dU}{dT} = -a_o\, k/q \tag{3.12}$$

From the theory of partial derivatives, the derivative of Eq. (3.12) can be obtained in terms of the ratio of the derivatives of Eq. (3.10) and (3.11); applying the usual mathematical formula leads to

$$\left(\frac{\partial V_b}{\partial T}\right)_I = \left(\frac{\partial \ell n I}{\partial T}\right)_{V_b} / \left(\frac{\partial \ell n I}{\partial V_b}\right)_T$$

$$= -(a_o/T) / (q/kT) = -a_o\, k/q \tag{3.13}$$

which is seen to agree, as it must, with the direct differentiation given in Eq. (3.12).

As an example of the order of magnitude of the quantities involved, we shall consider a particular case of a transistor having a charge of 10^{-6} coulombs per square centimeter of the base layer at room temperature and under conditions in which the current density is $1\,A/cm^2$. For silicon under these conditions the quantities needed for evaluating the derivatives are as follows:

$$q = 1.6 \times 10^{-19} \text{ coul} \tag{3.14a}$$

$$Q = 10^{-6} \text{ coul} / \text{cm}^2 \tag{3.14b}$$

$$n_i^2 = 2.2 \times 10^{20} \text{ cm}^{-6} \text{ at } 293°\text{K} \tag{3.14c}$$

$$n_{pn} = 3 \tag{3.14d}$$

$$V_g = 1.21 \text{ electron volts} \tag{3.14e}$$

$$D_n = 35 \text{ cm}^2/\text{sec} \tag{3.14f}$$

$$n_{\mu n} = -2.5 \tag{3.14g}$$

$$n_T = 1 - 2.5 + 3 = 1.5 \tag{3.14h}$$

$$I/A = 1 \text{ amp} / \text{cm}^2 \tag{3.14i}$$

$$U = kT / q = 293° / 11,600 = .025 \text{ volts} \tag{3,14j}$$

From these values the remaining single unknown V_b of Eq. (3.1) can be found:

$$V_b = U \ell n\, Q\, I/A\, q^2 n_i^2 D_n$$

$$= 0.025\, \ell n 10^{-6} / (1.6 \times 10^{-19})^2 \times 2.2 \times 10^{20} \times 35$$

$$= 0.025\, \ell n\, 0.52 \times 10^9 = 0.50 \text{ volts} \qquad (3.15)$$

Substituting these values in Eq. (3.8) leads to

$$a_o = [\,(1.21 - 0.5)/0.025] + 1.5$$

$$= 28.4 + 1.5 \cong 30 \qquad (3.16)$$

from which it is seen that in the operating range the current in effect depends upon a very high power of the temperature. Expressed as a temperature coefficient on the Kelvin scale, we obtain

$$a_T = a_o/T = 0.1°K^{-1} \qquad (3.17)$$

which signifies a change of approximately 10% in current for each degree increase in temperature on the Kelvin scale, or a factor of e increase in current for each 10°K.

The variation of base voltage with temperature at constant current, which may be used to measure temperature rise, has the value

$$\left(\frac{\partial V_b}{\partial T}\right)_I = -a_o k/q = -30/11,600$$

$$= -2.58 \text{ mv} / °K \qquad (3.18)$$

The condition for which the thermistor arrangement of Fig. 3.1 will exhibit a negative resistance can now be expressed in terms of

quantities familiar in connection with transistor design. A typical value for thermal resistance of a transistor is 10° per watt, leading to a value for the thermal conductance of

$$h_T = 0.1 \text{ watt}/°K \tag{3.19}$$

This leads to a relatively simple expression for the stability index s of Section 2:

$$s = a_T VI/h_T + (T - T_s)a_T \cong VI \tag{3.20}$$

For this particular example the transistor will exhibit a negative differential resistance on its V-I characteristic if the power exceeds 1 watt, and the temperature rise $T - T_s$ above a heat sink at temperature T_s exceeds 10°C.

It may be worthwhile discussing briefly the behavior of the transistor in the circuit of Fig. 3.1 in respect to the V-I characteristic which is obtained. If a fixed current is passed through the transistor, then a certain amount of heat will be generated. If the temperature rise caused by this heat is insufficient to make the transistor pass the required current through its base layer and across the collector junction, then the applied current will act to charge the collector capacitance and to raise the collector voltage. Under practical circumstances the voltage might well rise until avalanche effects occur. Neglecting avalanche effects for the moment, the rise in voltage at constant current will increase the power dissipated in the device and raise the temperature and increase the carrier current across the collector junction. By this mechanism the voltage will arrive at the value necessary to produce the current being

carried. This will be true for all values of current, and consequently although the transistor exhibits nearly zero high frequency conductance, it may give rise to a V-I characteristic having a negative resistance, as shown on Fig. 3.2. If the value of s is very much greater than unity, then this negative resistance characteristic will lie nearly along the constant power hyperbola, as discussed in connection with Fig. 2.2.

It should be pointed out that temperature gradients along the leads to base and emitter produce thermoelectric voltages which may appear across the emitter-base junction. Since thermoelectric coefficients for metals used for leads are of the order of 10^{-6} volts/degree and are largely compensated by the two leads being of the same metal, these thermoelectric voltages will in general have negligible influence.

In analytic studies of thermal effects over large factors of change in current, it is convenient to use the approximation

$$I(U) = I_s \exp [(U - U_s)a] \qquad (3.21)$$

where I_s is the current which would flow if the transistor were at the temperature U_s of the heat sink and U is the actual temperature, it being assumed that the emitter-base voltage is constant. This approximation involves the error of assuming that

$$\ell n I - \ell n I_s = (U - U_s)a \qquad (3.22)$$

where

$$a = a_o(U_s) / U_s \qquad (3.23)$$

This error is equivalent to using only the first term of a Taylor's expansion of $\ell n\, I$ in Eq. (3.5). The correct Taylor's expansion to two terms is

$$\ell n\, I - \ell n\, I_s = \{ (n/U_s) + [(V_g - V_b) / U_s^2] \} (U - U_s)$$

$$- \{ (n/U_s^2) + 2[(V_g - V_b) / U_s^3] \} (U - U_s)^2 / 2$$

$$= a(U - U_s) - [a - (n_T/2 U_s)] (U - U_s)^2 / U_s$$

$$= a(U - U_s) \{ 1 - [1 - (n_T/2aU_s)] (U - U_s) / U_s \} \qquad (3.24)$$

For many cases of interest the correction term is practically negligible, as can be seen as follows: $a U_s \doteq a_0$, which is approximately 30. Hence the $n_T/2a U_s$ term is small compared to unity. For cases of interest $a(U - U_s)$ is approximately unity, so that

$$1 \doteq a(U - U_s) \doteq a U_s (U - U_s) / U_s$$

$$= a_0 (U - U_s) / U_s \qquad (3.25)$$

Hence $(U - U_s) / U_s \doteq 1 / a_0$. Thus the correction to the first term of the Taylor's expansion is only a few percent, and it is satisfactory to use (3.21).

For large variations in temperature it is sometimes necessary to take into account changes in a_0. These are usually small, provided the current density is kept constant. This conclusion may be reached by writing a_0 in the form

$$a_0 = [(V_g - V_b) / U] + n_T \qquad (3.26)$$

where from equation (3.1) we obtain

$$(V_g - V_b) / U = -\log IQ / A q^2 D_n n_i^2 \exp(V_g / U)$$

$$= -\log I + n_T \log U + \text{const.} \qquad (3.27)$$

In this expression V_b is the forward voltage across the emitter junction. For the case where $(V_g - V_b) / U$ is the order of 30, changes of a factor of two in absolute temperature will change $(V_g - V_b) / U$ by less than 10% from the $n_T \log U$ term.

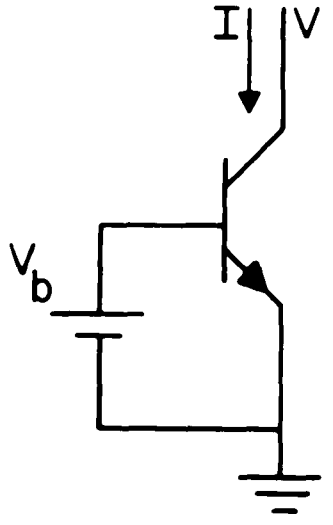


Figure 3.1

The transistor as a thermistor.

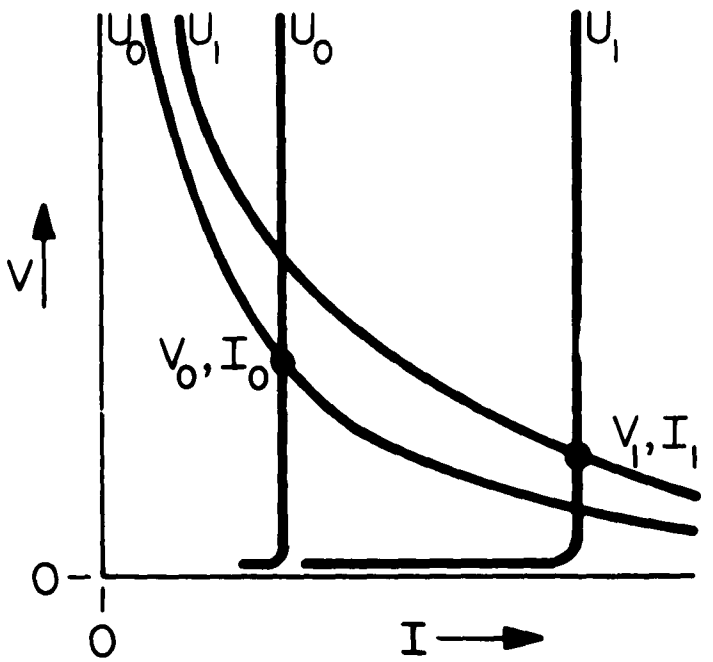


Figure 3.2

Negative resistance V-I characteristic for saturated collector current.

4. INSTABILITY OF TWO PARALLEL THERMISTORS

If two thermistors are connected in parallel and sufficient current is passed through them to produce a negative resistance condition, then a phenomenon known as "current-hogging" will occur. This condition corresponds to an instability building up, in which one thermistor carries more current than the other and consequently becomes warmer and drops its resistance. This inequality then tends to build up until finally essentially all of the current is going through one of the thermistors and the other one is relatively cool and carries little current.

The instability condition depends not only upon the characteristics of the individual thermistors but also upon the ability of heat exchange to take place between the two of them. This situation is represented schematically in Fig. 4.1. In this figure the thermal conductances of each thermistor to a heat sink are represented by the heat conductance h. On the Figure heat conductances are represented by wavy lines and electrical conductances by straight lines. There is also assumed to be a mutual heat conductance m between the two thermistors. Evidently, if the two thermistors were actually connected mechanically into such intimate contact that there could be no temperature difference between them, then it would be impossible for one of them to become heated in respect to the other and carry substantially all of the current.

The problem of this section is to derive the analytical conditions determining the stability of the pair of parallel thermistors. This stability problem is very similar to the problem of hot spot generation in a power transistor, in which one portion of the power transistor takes over all of the current and becomes warm, leaving the rest of the transistor relatively cool.

Fig. 4.1 represents the condition we shall study in schematic form. We assume that for each thermistor by itself a point on the V-I characteristic corresponds to values V_0 and I_0. The disturbed condition for the parallel circuit is represented on the figure, with currents i_1 and i_2 representing the disturbances in current through each of the two thermistors. In keeping with Eq. (2.6) we have two equations for the two thermistors:

$$K\dot{u}_1 = V_0 i_1 + I_0 v - h u_1 - m(u_1 - u_2) \tag{4.1}$$

$$K\dot{u}_2 = V_0 i_2 + I_0 v - h u_2 - m(u_2 - u_1) \tag{4.2}$$

In these equations the terms involving the mutual thermal conductance m represent a heat flow between the two thermistors.

In keeping with Eq. (2.5) the disturbances in current through the thermistors are given by

$$i_1 = g v + I_0 a u_1 \tag{4.3}$$

$$i_2 = g v + I_0 a u_2 \tag{4.4}$$

The four preceding equations involve five unknown quantities, the two disturbances in current, the two disturbances in temperature, and the disturbance v in voltage across the parallel pair. If, for example, the disturbance in voltage were specified, then these would represent four equations for the four unknowns. In the linear range which we study, the problem of instability between the two transistors is fortunately independent of the form of the applied voltage wave. This independence may be seen by introducing disturbance variables which depend upon the difference in behavior of the two thermistors. Thus we introduce two new variables, as follows:

$$u_1 - u_2 = u_3 \qquad , \qquad i_1 - i_2 = i_3 \tag{4.5}$$

In terms of these variables, the first four equations may be reduced to a simpler form simply by subtracting Eq. (4.2) from (4.1) and Eq. (4.4) from (4.3). The results are

$$K\dot{u}_3 = V_0 i_3 - (h + 2m)\, u_3 \tag{4.6}$$

$$i_3 = I_0 a u_3 \tag{4.7}$$

It is seen that these two equations are identical in form to Equations (2.5) and (2.6), which describe a single thermistor for the case of $v = 0$; this last condition implying connection to a zero impedance external circuit. Fig. 4.2 shows the equivalent circuit corresponding to Equations (4.6) and (4.7). It is seen that the thermistor involved

behaves as if it had, in addition to the thermal conductance h, an additional thermal conductance of 2m. This corresponds to the fact that the unstable behavior involves one thermistor heating up and the other one cooling down by an equal amount, so that the conductance of heat through the mutual conductance m is twice as great as if one side were held at a fixed temperature. This is why the conductance 2m appears in Eq. (4.6).

It should particularly be noted that the disturbance u_3, i_3 occurs independently of the impedance of the external power source so that no changes in this source can influence the instability associated with u_3 and i_3. We shall return to this point after introducing u_4 and i_4 below.

The stability condition corresponding to the equivalent circuit of Fig. 4.2 is evidently

$$s_3 = a V_0 I_0 / (h + 2m) \qquad (4.8)$$

As noted above, Eq. (4.7) corresponds to the situation in which there is no applied voltage, and thus to a situation in which the equivalent circuit of Fig. 4.2 is fed from a zero impedance source. As was discussed in connection with Fig. 2.3, instability occurs for a zero-impedance terminated condition if the stability index becomes greater than unity or if the equivalent circuit shows a dc negative conductance. This condition of instability for Fig. 4.2 is seen to correspond to a higher value of s_3 than would occur if there were no mutual thermal conductance m present.

For completeness, we note that if i_3 is eliminated from (4.6) and (4.7) we obtain

$$K\dot{u}_3 = VIa\,u_3 - (h + 2m)u_3 \tag{4.9}$$

This shows that the temperature difference u_3 obeys the equation

$$[K / (h + 2m)]\,\dot{u}_3 = (s_3 - 1)u_3 \tag{4.10}$$

which corresponds to an exponential build-up of temperature difference if s_3 is greater than unity. Under these conditions u_3 varies as

$$u_3 = \text{const} \exp(t / \tau_{s3}) \tag{4.11}$$

where the characteristic build-up time τ_{s3} is given by

$$\tau_{s3} = \tau_s / (s_3 - 1) \tag{4.12}$$

where τ_s is the thermal relaxation time of the equivalent circuit of Fig. 4.2, neglecting any electrical effects:

$$\tau_s = K / (h + 2m) \tag{4.13}$$

Although the thermal instability term represented by u_3 is the important term for our consideration, it is worthwhile to point out that in addition to this solution there is another solution to the first four equations describing the pair of parallel thermistors of this section. These other conditions can be expressed in terms of two additional variables u_4 and i_4, as follows:

$$u_1 + u_2 = 2u_4 \quad , \quad i_1 + i_2 = 2i_4 \tag{4.14}$$

In terms of these quantities, Equations (4.1) to (4.4) can be combined into two equations:

$$K\dot{u}_4 = V_0 i_4 + I_0 v - h u_4 \tag{4.15}$$

$$i_4 = g v + I_0 a u_4 \tag{4.16}$$

These equations are seen to be identical with the behavior of a single isolated thermistor. Thus if the instability does not build up, the pair of thermistors simply act like two thermistors in parallel, and each one behaves in precisely the same way as it would in the absence of the other one.

If the external circuit has such high impedance that no instability occurs with i_4 and u_4, then it is still possible for i_3 and u_3 to build up if $s_3 > 1$. In this case since $i_4 = 0$

$$i_4 = 0 = i_1 + i_2 \tag{4.17a}$$

$$u_4 = 0 = u_1 + u_2 \tag{4.17b}$$

it follows that

$$i_1 = -i_2 = i_3 / 2 \tag{4.18a}$$

$$u_1 = -u_2 = u_3 / 2 \tag{4.18b}$$

so that the build-up does not affect the external circuit behavior until the disturbance exceeds the amplitude where the linear approximations are valid.

The condition $s_3 = 1$ is a second example of self-balance. (Self-balance was discussed in connection with Eq. (2.13).) and is closely related to the problem of instability of the distributed structure discussed in Section 5.) For the parallel thermistor case, when $s_3 = 0$, a slight increase of current in thermistor 1 and equal decrease in thermistor 2 results in a disturbance of power which exactly balances the disturbances in heat flow--again a self-balance situation. So far as the thermal properties are concerned, the stability condition does not involve the heat capacitances but only the heat conductances. This is because for $s_3 = 1$ neither build-up nor decay occurs, so that no temperatures change; consequently, no power flow to heat capacitances is involved. An increase in general power level, making $s_3 > 1$, will destroy the self-balance and will result in more electrical power-change than needed for the changed thermal energy flows. The rate of build-up will then depend on the heat capacitances as shown by Eq. (4.11) and (4.13).

It should be noted that our discussion has dealt analytically only with the small signal disturbance. It is difficult to extend the treatment analytically to large signals, but graphical methods can be used to visualize the results which will occur. The important criterion is generally the one which governs the initiation of the instability. After this, in general, it will proceed to some physically recognizable limit, such as the one discussed earlier in this section, indicating that the current will predominantly flow through one

of the thermistors. In the case of a power transistor, other adverse limitations may occur, such as concentration of the current to such a degree that the device is destroyed by local heating.

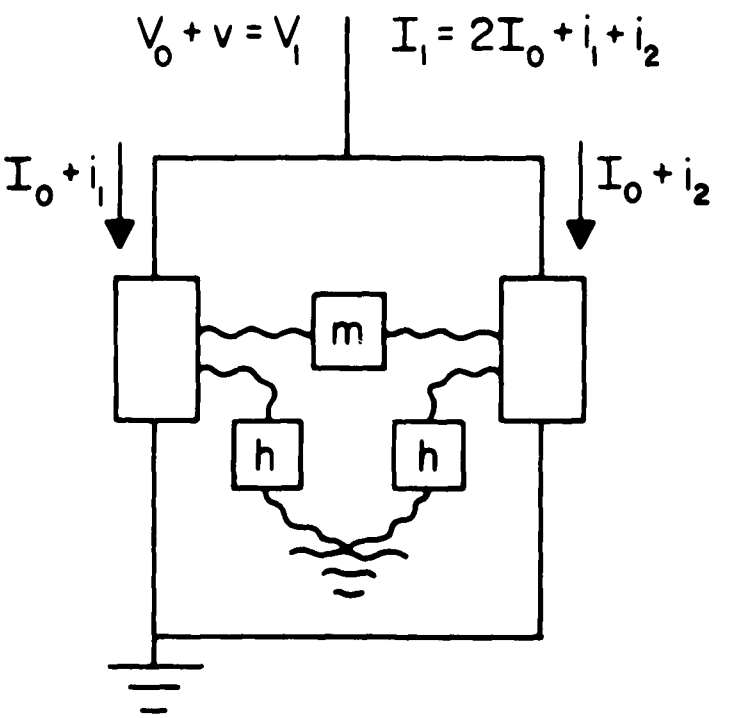


Figure 4.1

Two thermistors in parallel.

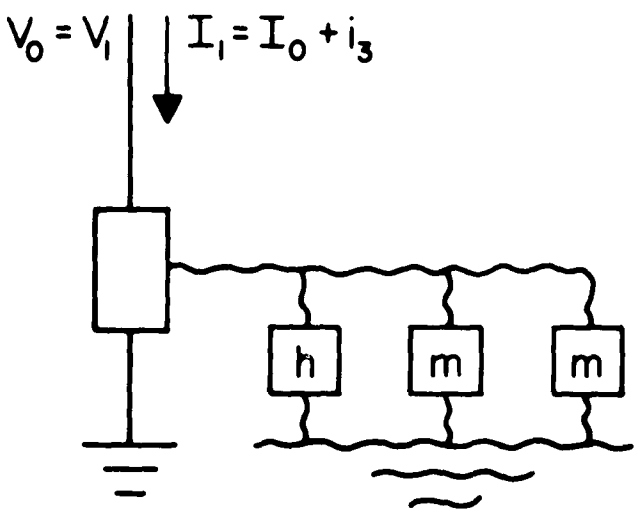


Figure 4.2

Equivalent circuit for the unstable behavior of two thermistors.

5. THE DISTRIBUTED THERMISTOR OR TRANSISTOR STRUCTURE

5A. Introduction

The problem of the distributed case of a transistor or thermistor in which one portion becomes heated and steals the current from the remainder is qualitatively similar to the two thermistor problem of Section 4. The decisively important quantity is essentially the stability index expressed as

$$s = a_T \Delta T_0 = a \Delta U_0 \tag{5.1}$$

where $\Delta T_0 = q \Delta U_0 / k$ is the temperature rise in degrees Kelvin of the active portion of the device above the heat sink at temperature T_s and a and a_T are the coefficients of current increase at constant voltage per volt and per °K as discussed in Sections 2 and 3.

The roles of m, the mutual thermal conductance and h, the thermal conductance to the heat sink, discussed in Section 4, have counterparts in the distributed system. The detailed behavior in any particular case will be complicated by the boundary conditions and in some cases will be nearly impossible to solve analytically. However, many of the main features can be understood in terms of simplified examples that can be treated in detail.

In the following two subsections two cases will be discussed. In Subsection 5B the steady state condition corresponds to the condition

V_0, $2I_0$ for the two thermistors of Section 4. The entire area where electrical power is converted to heat is at a uniform temperature. The small disturbance of the two thermistors corresponding to u_3 and i_3 is found to be represented by a continuous sinusoidal small disturbance in temperature over the area. Stability conditions are worked out in terms of possible small disturbances.

Subsection 5C covers a case in which the steady state produces a temperature distribution which varies over area of heat generation. This leads to more complex analytical problems. The new aspects are treated in detail. Conclusions about some parts which are similar to those of subsection 5B but analytically more complex are reached on the basis of logical extensions of the results of subsection 5B.

5B. Uniform Steady State Temperature

Fig. 5.1 illustrates the structure considered. It is assumed that the electrical energy is converted to heat in a very thin thermistor layer lying on the surface of a block of height Y and width X. We shall initially assume that there is no dependence upon Z of the effects investigated. The surfaces at $x = 0$ and $x = X$ are assumed to be thermally insulated and the surface at $y = 0$ is an ideal heat sink at a temperature T_s.

The block is assumed to have uniform thermal properties described by a thermal conductance c in Watts / cm volt or amp / cm for temperature expressed as volts in U units or c_T where

$$c_T = kc / q \quad ; \quad q / k = 11,600 \text{ volt} / {}^\circ K \tag{5.2}$$

c_T is in Watts / cm °K for T in °K. The thermal diffusion constant D_T is in cm^2 / sec and does not depend on the unit chosen for temperature.

This model may also be applied to a transistor structure. The model is particularly applicable to a diffused transistor. For this application the thermistor layer comprises the emitter, base and space charge region of the collector, as represented on Fig. 5.2. The heat is generated in this region, which is usually relatively thin compared to the collector body as a whole. The electrical contacts are differently arranged from those of Fig. 5.1. However, as discussed in Section 3, for a fixed emitter-base voltage, the transistor has a current which depends chiefly upon the temperature of the emitter-base region.

For the models of Figs. 5.1 and 5.2, it will be supposed that the cur-

rent density along the y-axis per unit area of top surface is

$$I_a = F_a(V, U) \tag{5.3}$$

where the subscript a signifies per unit area. For the transistor, F_a is practically independent of V the emitter-collector voltage for the case treated here, in which saturation current is drawn across the collector junction; the emitter-base voltage is regarded as a fixed constant in this section and is, therefore, omitted from the expressions used.

Two simplifications should be mentioned here. In the transistor the current density actually has a complex structure dependent on the interdigitation of the emitter and base fingers. As will be discussed in Section 6, the fine scale fluctuations in current may be replaced by uniform average values provided a correction is made in the effective thermal resistance. It is also assumed that variations in temperature over the top surface do not produce significant differences in emitter-base voltage through thermoelectric effects; this point was previously discussed in Section 3, where it was pointed out that thermoelectric voltages generated in the metal leads and contacting bodies of the emitter and base structures were generally negligible.

Considering Fig. 5.1 or 5.2, we assume that the steady state current density is represented by a uniform average value I_{a0} and that the applied voltage is V_0, so that the power density is

$$P_{a0} = I_{a0} V_0 = c_T \Delta T_0 / Y \tag{5.4}$$

where $\Delta T_0 / Y$ is the temperature gradient from the heat sink to thermistor

layer. In terms of this ΔT_0 we can introduce a stability index s by

$$s = a_T \Delta T_0 = a_T Y P_{a0} / c_T = a Y P_{a0} / c \tag{5.5}$$

where, as discussed in Section 2 for a and h, the ratio $a/c = a_T / c_T$ is independent of the temperature unit involved.

Assuming uniform power supply on the surface y = Y, no heat flow across x = 0 or x = X, and uniform temperature T_s at the heat sink at y = 0, the steady state distribution of Fig. 5.1 is

$$T_0(x, y, t) = q U_0(x, y, t) / k = T_s + (P_{a0} y / c) \tag{5.6}$$

in other words a linear temperature gradient in the y direction. This corresponds to the steady state V_0, $2 I_0$ distribution for the two thermistors of Section 4.

The disturbance of Section 4 corresponds to introducing an unknown disturbance of temperature u(x, y, t) which must satisfy the boundary conditions and differential equations, and analyzing its behavior. The boundary conditions appropriate to the problem are as follows:

$$\text{at } y = 0; \qquad u(x, 0, t) = 0 \tag{5.7}$$

$$\text{at } y = Y; \qquad c \partial u / \partial y = a P_{a0} u(x, Y, t) + (V_0 g_a + I_{a0}) v(t) \tag{5.8}$$

$$\text{at } x = 0 \text{ or } X; \qquad \partial u / \partial x = 0 \tag{5.9}$$

The first condition is required by the assumption that no temperature change can be produced on the perfect heat sink at y = 0. The second condition states that the transient disturbance in power in the thermistor layer is

carried away by thermal conduction. The last two conditions imply no heat flow through the surfaces x = 0 and x = X.

In the body the thermal diffusion equation must hold:

$$\partial u / \partial t = D_T (\partial^2 u / \partial x^2 + \partial^2 u / \partial y^2) \tag{5.10}$$

A set of functions which satisfy (5.7) and (5.9) and (5.10) are

$$u_n = \cos[n\pi(x/X) - n(\pi/2)] \sinh b_n y \exp \nu_n t \tag{5.11}$$

$$\nu_n = D_T[-(n\pi/X)^2 + b_n^2] \tag{5.12}$$

where n = 0, 1, 2, etc. The term for n = 0 corresponds to a disturbance which is uniform over the area of the structure. It corresponds to the situation of Section 4 described by u_4, i_4 in which the two thermistors behave alike. It represents the response of the structure as a whole to changing voltages and currents and does not contribute directly to the instabilities of interest. The n=1 term is of special importance discussed below.

In order to satisfy (5.8), all terms with n greater than zero must satisfy

$$c b_n \cosh b_n Y = a P_{a0} \sinh b_n Y \tag{5.13}$$

or

$$b_n Y \coth b_n Y = \theta \coth \theta = a P_{a0} Y / c = s \tag{5.14}$$

For a given value of s, this equation is satisfied for an infinite set of θ values, as may be seen from Fig. 5.3. (For complex values of θ, $\theta \coth \theta$ is also complex and Eq. (5.14) cannot be satisfied.) For real values of θ, $\theta \coth \theta$ is > 1, so that real values of θ can occur only if $s > 1$. For all values of s an infinite series of imaginary values of θ are possible. These imaginary values correspond to imaginary values of the b_n quantities and

thus give rise to terms which depend on y through ordinary sine functions rather than hyperbolic sines. These terms are needed to expand an arbitrary disturbance u(x, y, t) in the space of the device, but they all have negative values of ν_n and decay with time. They are therefore not responsible for causing instabilities. (Further aspects of the various terms can be investigated and formalized by the usual procedures for investigating sets of eigenfunctions.)[3]

The critical term in regard to stability is that with n = 1. As s increases above unity so that θ and $b_n = \theta/Y$ become real and greater than zero, the first ν_n to become positive and give an exponential build-up in time is readily seen from Eq. (5.12) to be ν_1. Thus the critical condition for stability is that $\nu_1 = 0$ and this leads to

$$b_1 = \theta / Y = \pi / X \tag{5.15}$$

or

$$\theta = \pi Y / X \tag{5.16}$$

so that the critical condition for s becomes

$$aP_{a0} Y/c = s = \theta \coth \theta = (\pi Y / X) \coth (\pi Y / X) \tag{5.17}$$

We shall next give some physical interpretation to this stability condition.

The spatial part of the n = 1 term involves x through the factor $\sin (\pi x / X)$. It represents a disturbance in which equal and opposite temperature changes occur across the center line x = 0.

Consider first the case of $Y / X << 1$. Then θ approaches zero so that s approaches unity. Thus the stability condition is exactly as

---

[3] See, for example, Courant-Hilbert, Methoden der Mathematischen Physik, Dover Publications, New York, 1947.

if the structure consisted of two separate thermistors in parallel, each with s = 1 and no mutual thermal conductance. The interpretation is that for $Y << X$; thermal conduction to the heat sink is so much more important than lateral conduction that the latter can be neglected.

For $Y/X >> 1$, the situation is very different. For this case coth θ approaches unity and we have

$$s = aP_{a0} Y/c - \pi Y/X \tag{5.18}$$

This leads to

$$aP_{a0}(X/\pi)/c = 1 \tag{5.19}$$

This is the equation that would be obtained for a simple thermistor situation in which the heat flow path is X/π long. In effect, the structure is stabilized by conduction in the x-direction. Under these conditions the quantity obtained by multiplying a by the temperature rise above the heat sink at temperature T_s is

$$aP_{a0} Y/c = (aP_{a0} X/\pi c)(\pi Y/X) = \pi Y/X \tag{5.20}$$

This quantity is much greater than unity. This is an example of a situation in which the entire thermistor unit may have a negative resistance due to heating effects, and if connected in parallel with a similar unit would result in instability. Internally, however, it will be stable because of lateral heat conduction along the x-axis provided s is less than $\pi Y/X$.

As a final observation on distributed systems, it should be pointed out that a simple physical interpretation in terms of the self-balance concept of Sections 2 and 4 can be given to the condition

$$\nu_1 = 0 \tag{5.21}$$

that divides the stable from the unstable situation. The disturbance $u(x, y, t)$ for $\nu_1 = 0$ is one which does not change in time and can thus be represented as $u(x, y, 0)$. Evidently $u(x, y, 0)$ is such a disturbance that power which flows away as thermal conduction due to the temperature disturbance is exactly balanced by the electrical power generated by changes in current at constant voltage produced by the effect of $au(x, y, 0)$ on the current density. In other words, this is again the condition of self-balance discussed in Sections 2 and 4. It is closely related to the parallel thermistor case of Section 4; in Section 4 the instability term arose from disturbances which were equal and opposite for the two thermistors. For the distributed case, the $\cos(\pi x / X)$ term in $u_1(x, y, t)$ represents a disturbance in which changes in current and temperature for x in the range $-(X/2) < x < 0$ are equal and opposite to changes in the range $0 < x < (X/2)$. Again, as for the two thermistors in parallel of Section 4, the instability does not involve the impedance of the external circuit since no net change in current occurs and the disturbance v in voltage from equation (5.8) does not enter into equation (5.14) for the boundary condition at $y = Y$; this corresponds to the absence of v

from the terms involving i_3 and u_3 in Section 4.

Since for $\nu_1 = 0$, the self-balance is perfect, no power flow into heat capacitance is required and consequently the heat capacitance does not enter the equations for stability.

It is physically evident that the self-balance condition for the distributed structure represents the marginal condition of stability. In contrast, if an increase in temperature u(x, y, 0) were to produce an extra power flow which, flowing out through the thermal resistance, produced an increase in temperature greater than u(x, y, 0), then it is evident that thermal energy would accumulate and the temperature rise would increase with the passage of time. Thus one procedure for determining stability limits is to find temperature disturbances which lead to self-balance. In general, any increase in power above these levels will lead to instability.

Once instability occurs, the rate of build-up does depend on heat capacity. For the critical term ν_1, the build-up time constant is

$$\nu_1 = D_T [b_1^2 - (\pi / X)^2] = D_T [(\theta / Y)^2 - (\pi / X)^2]$$

$$= [\theta^2 - (\pi Y / X)^2] \div (Y^2 / D_T) \qquad (5.22)$$

In this expression the term Y^2 / D_T is a thermal diffusion time for heat to diffuse the distance Y. Since

$$D_T = c_T / \text{(heat capacity per unit volume)} \qquad (5.23)$$

it contains the effect of heat capacitance. The build-up constant may be

much shorter or longer than this time, depending on how much larger θ is than $(\pi Y / X)$. It is evident that for power levels much larger than the stable value, the rate of build-up increases rapidly with increasing power. For values of s much greater than unity, θ and s are nearly equal and both are proportional to the power density P_{a0} in watts/cm^2. Thus the time to build-up should vary inversely as power squared in this range.

5C. Varying Steady State Temperature; Heat Sinks at the Ends

Another example of a distributed structure is shown in Fig. 5.4. In this case the heat sinks are located at the ends of the structure on the planes at $x = -X/2$ and $x = X/2$. It is assumed that the structure is thin compared to its length so that the temperature may to a satisfactory approximation be regarded as independent of y. This structure will show negative resistance when the appropriate stability index s becomes high enough. If driven by a constant current source, it will acquire an internal instability in which the portions at large and small z values act like two parallel thermistors.

This case is more complex than that of Subsection 5B because in the steady state condition the temperature and power densities are functions of x. For a small disturbance $u(x, z, t)$ this situation leads to a differential equation which does not have constant coefficients, but instead coefficients that are functions of x. It is fortunately not necessary to solve these equations or even to write them down in order to reach conclusions about thermal instabilities. These can be seen to arise when negative resistance occurs, by reasoning on the basis of the considerations of Subsection 5B. The results obtained for this somewhat simplified case throw light on how an actual power transistor structure may deviate from the example of Subsection 5A.

For the structure of Fig. 5.4, we define a heat conductance h as follows: (Since we are here dealing with only a steady state case the

subscript "0" used to distinguish between steady state and subscript "1" for a disturbed state can be omitted to simplify the equations.) Suppose power is uniformly applied over the x-z plane with power density P_a watts/cm^2 produced by current density I_a and voltage V; then the total power XZP_a is

$$VI = P = XZP_a \tag{5.24}$$

and the highest temperature rise occurs at the middle line x = 0, which is farthest from the heat sink; this temperature rise is

$$U = U_s = P/h = XZP_a/h \tag{5.25}$$

where h is an effective heat conductance. The value of this heat conductance may be calculated simply as follows: Half the power flows out to the right hand heat sink at x = X/2. Furthermore, on the average, this power originates at X/4 from the heat sink. The thermal conductance for a distance X/4 is 4c YZ/X, where c is the thermal conductivity in U-units of Eq. (5.2). Two such thermal conductances act in parallel. Hence the total effective thermal conductance is

$$h = 8cYZ/X \tag{5.26}$$

In terms of this value of h, the stability index s is defined in keeping with Eq. (2.13) in terms of the total power P:

$$s \equiv aP/h = aXP/8cYZ = aXVI/8cYZ \tag{5.27}$$

where V is the voltage and I the total current.

The current depends upon x through the dependence of the temperature U(x). As shown in Eq. (3.13), for a transistor-like device

the current density can be expressed as

$$I_a(x) = I_{as} \exp (U(x) - U_s)a \tag{5.28}$$

where U_s is the temperature of the heat sink at $x = \pm X/2$ and I_{as} is the current density at temperature U_s.

As will be shown below, negative resistance occurs when $s = 1$, which requires a power, defined as P_s:

$$P_s = 8c\,Y\,Z/a\,X \tag{5.29}$$

It is convenient to define a voltage and current in terms of P_s and I_{as} as follows:

$$P_s = V_s I_s = 8c\,Y\,Z/a\,X \tag{5.30}$$

$$I_s = I_{as} X Z \tag{5.31}$$

$$V_s = P_s/I_s = 8c\,Y/I_{as}\,a\,X^2 \tag{5.32}$$

Under operating conditions the temperature is not uniform but rises to a maximum $U(0)$ at $x = 0$. We introduce a variable θ related to this maximum by the expression

$$(U(0) - U_s)a = 2\,\ell n \cosh \theta \tag{5.33}$$

The current at $x = 0$ is then found from Eq. (5.28) to be larger than the value at the heat sink by the factor

$$I_a(0)/I_{as} = \exp 2\,\ell n \cosh \theta = \cosh^2 \theta \tag{5.34}$$

The value of θ corresponding to a given power P may be obtained by solving the equation

$$P = P_s\,\theta \tanh \theta \tag{5.35}$$

This relationship, together with the corresponding expressions for voltage and current

$$I = I_s \, (\sinh 2\theta) / 2\theta \tag{5.36}$$

$$V = V_s \, \theta^2 / \cosh^2 \theta \tag{5.37}$$

can be derived from the heat flow equation and equation (5.28). It should be noted that θ increases monotonically with P and I; however, V rises to a maximum and then falls. This maximum occurs when the derivative of $\theta^2/\cosh^2\theta$ vanishes and leads to

$$\theta \tanh \theta = 1 \quad , \quad \theta = 1.2 \tag{5.38}$$

so that $P = P_s$ and

$$s = P/P_s = a P X / 8 c Y Z = 1 \tag{5.39}$$

when negative resistance occurs. The corresponding values of V and I are

$$V = 0.44 \; V_s \tag{5.40}$$

$$I = 2.27 \; I_s \tag{5.41}$$

The above results can be derived by writing the heat flow equation for a thin slab so as to express the divergence of heat flow as the rate of generation per unit area:

$$\begin{aligned} -c Y \; [(\partial^2 U/\partial x^2) + (\partial^2 U/\partial y^2)] \\ = V I_{as} \exp \; [U(x) - U_s] \; a \\ = P_{as} \exp \; [W_s - W(x)] \end{aligned} \tag{5.42}$$

where cY is the heat conductance per square, and P_{as} the power density at the heat sink depends only on applied power through V. It is assumed

that the power generation terms depend only on x so that the partial derivative in respect to x becomes a total derivative and that for y vanishes. The variable W is defined as

$$W(x) \equiv [\, U(0) - U(x) \,]\, a \tag{5.43}$$

and is the logarithm of the ratio of current (or power density) at the maximum (at x = 0) to its value at x. Evidently at the heat sink the value of W is

$$W_s \equiv W(X/2) = [\, U(0) - U(X/2) \,]\, a = [\, U(0) - U_s \,]\, a \tag{5.44}$$

Equation (5.42) takes a simpler form by introducing a reference length

$$L_\theta = (c\, Y / a\, P_{as} \exp W_s)^{1/2} \tag{5.45}$$

and a reduced distance variable

$$x_\theta = x / L_\theta \tag{5.46}$$

The reference length L_θ is seen to be the distance in which the quantity aU(x) would decrease by 0.5 from its maximum at x = 0 if the power density level was uniformly $P_{as} \exp W_s$. In terms of x_θ Eq. (5.42) then becomes

$$d^2W/d^2x_\theta = \exp(-W) \tag{5.47}$$

The function $W(x_\theta)$ which satisfies this equation W = 0 at x = 0, as required by Eq. (5.43), is found to be

$$W = 2\, \ell n \cosh x_\theta / 2^{1/2} \tag{5.48}$$

as may be verified by differentiating twice.

The quantity θ introduced above in equation (5.33) is the argument of the cosh function at x = X/2:

$$W_s = 2\ \ell n\ \cosh\ \theta \tag{5.49}$$

$$\theta\ = (X/2L_\theta)/2^{1/2} \tag{5.50}$$

In terms of these quantities the expression for total power P can be put into the form of Eq. (5.35) as follows: The power P is equal to the heat flow out the two ends at $x = \pm X/2$:

$$P\ = -2\,c\,Y\,Z\,(dU/dx)_{X/2}$$

$$= (2\,c\,Y\,Z/a\,L_\theta)(dW/dx_\theta) \tag{5.51}$$

In this expression the derivative of W is evaluated at

$$x_\theta = x_{\theta s}\ \equiv X/2\,L_\theta = \theta\,2^{1/2} \tag{5.52}$$

and has the value

$$dW/dx_\theta = 2^{1/2}\ \tanh\ \theta \tag{5.53}$$

Introducing (5.53) for the derivative and L_θ from (5.50) into Eq. (5.51) leads to the desired equation (5.35) for P as a function of θ.

The expression for V is obtained by solving Eq. (5.45), the definition of L_θ, for $V(\theta) = P_{as}\ (\theta)\ /\ I_{as}$ where the functional dependence upon θ emphasized that I_{as} is evaluated at the heat sink and does not depend on θ or P. The result is

$$V\ = P_{as}/I_{as}\ = c\,Y/I_{as}\ a\,L_\theta^{\,2}\ \exp\ W_s$$

$$= (8\,c\,Y/I_{as}\ a\,X^2)\ \theta^2\ \cosh^2\ \theta\ = V_s\ \theta^2/\cosh^2\theta \tag{5.54}$$

From this expression for V and Eq. (5.35) for P, expression (5.36) is readily obtained for I from $I = P/V$ and $P_s = V_s I_s$.

The simple stability criterion of $s = 1$ obtained in this example is the result of several compensating effects. One of these is that as the power rises, the heat generation shifts more towards $x = 0$, thus raising the effective thermal resistance. At the center line $x = 0$ the temperature rise corresponds to $[U(0) - U_s]\, a = W_s = 2 \ell n \cosh \theta = 1.19$. Since this value differs from unity, this temperature rise does not correspond to the simple $s = 1$ condition of Section 2 and represents a compromise between non-uniform heating and the shift of power distribution discussed above. The power density at the center is $\exp W_s = 3.25$ times that at the heat sink.

If the strip is relatively wide in the z-direction, it is evident that considerations like those for the structure of Subsection 5A apply. One part of the strip may steal current from the other part. Eq. (5.42) may be used to treat this case, using a disturbance of the form $u(x) \cos (\pi z/Z)$. For such disturbances the line $z = Z/2$ will not be disturbed and will act like a heat sink maintained at a temperature $U(x)$ corresponding to the steady state. It is thus evident that unless Z is much larger than X, important lateral heat conductance effects in the z-direction will occur, and the condition $s = 1.0$ will have to be significantly exceeded before lateral instability occurs.

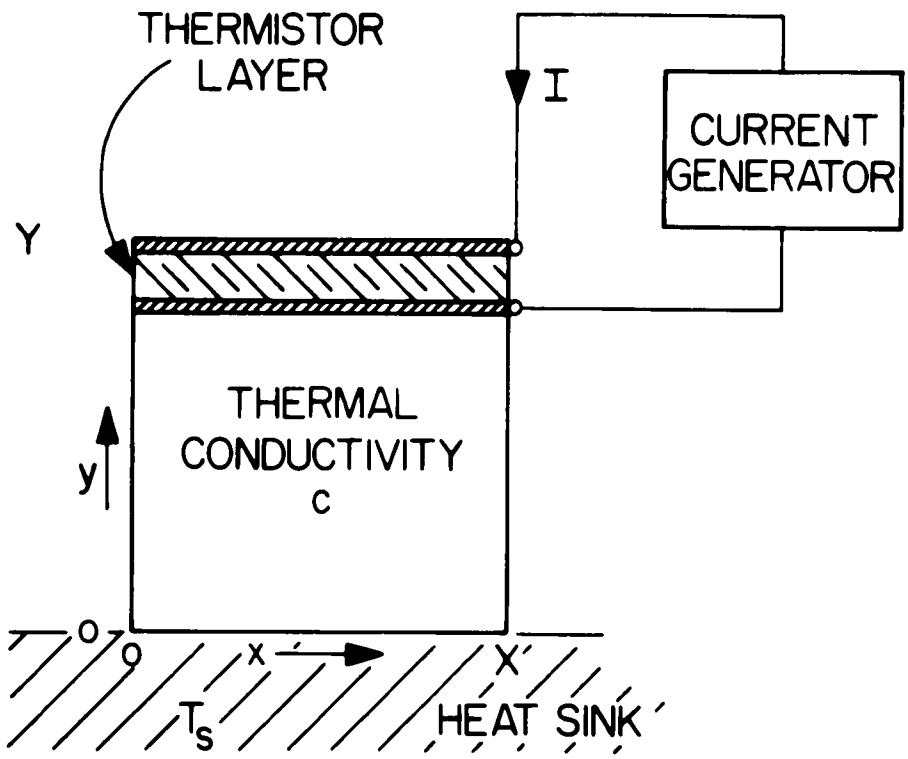


Figure 5.1

Distributed thermistor model.

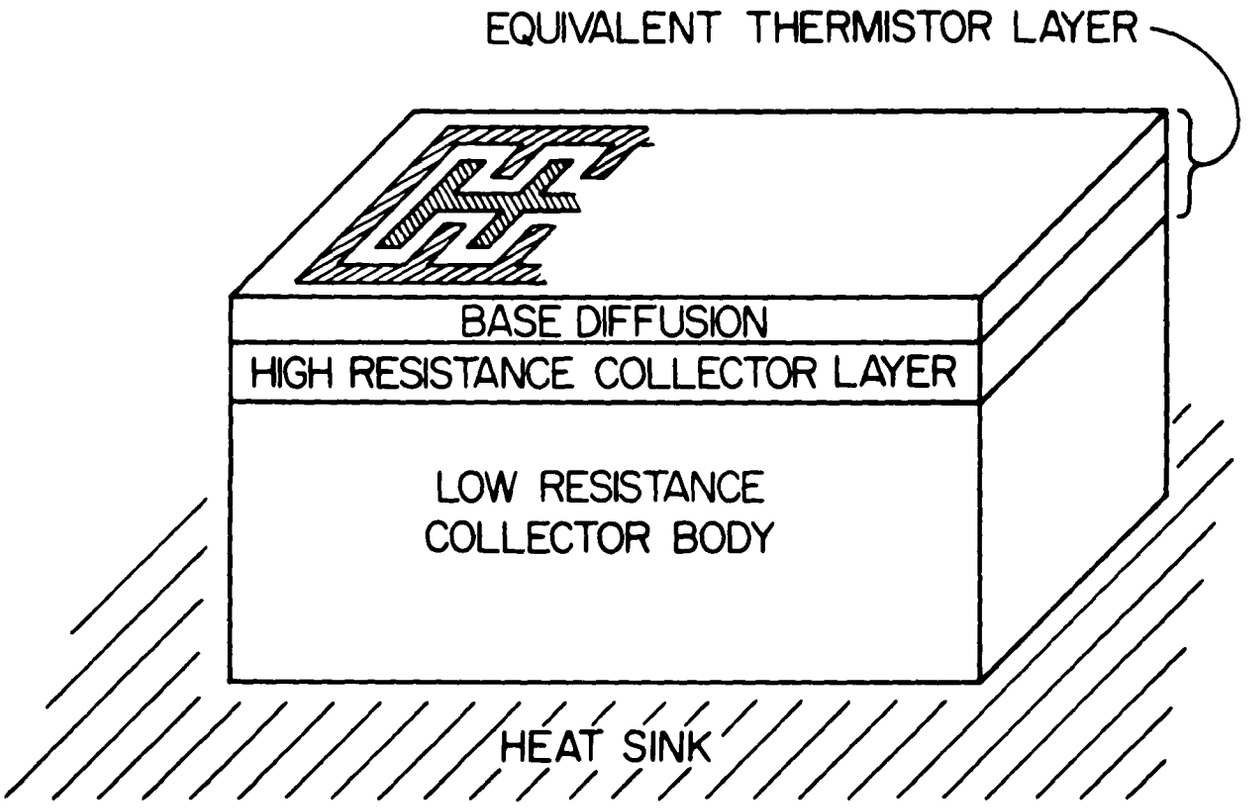
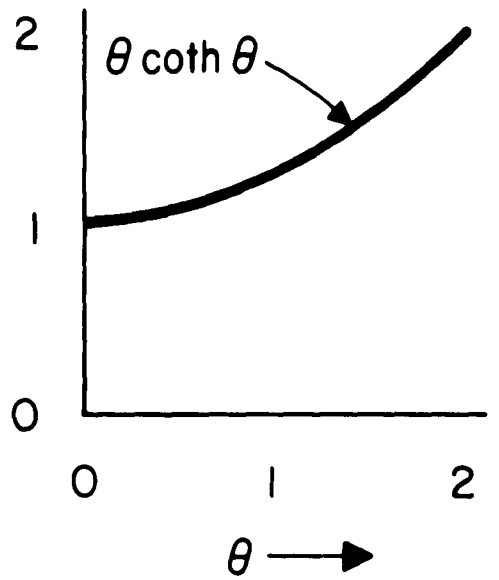


Figure 5.2

Transistor as distributed thermistor



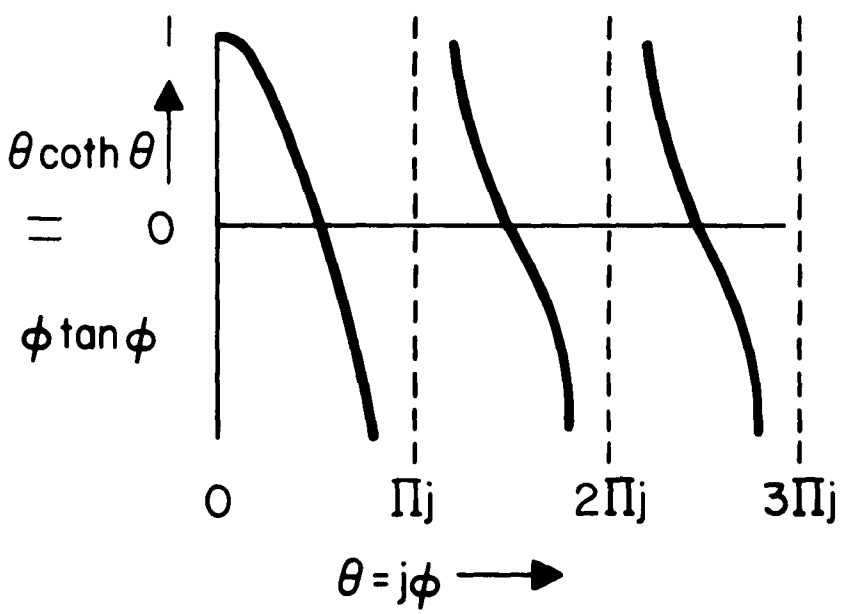


Figure 5.3

The value of θ coth θ for pure real and pure imaginary values of θ.

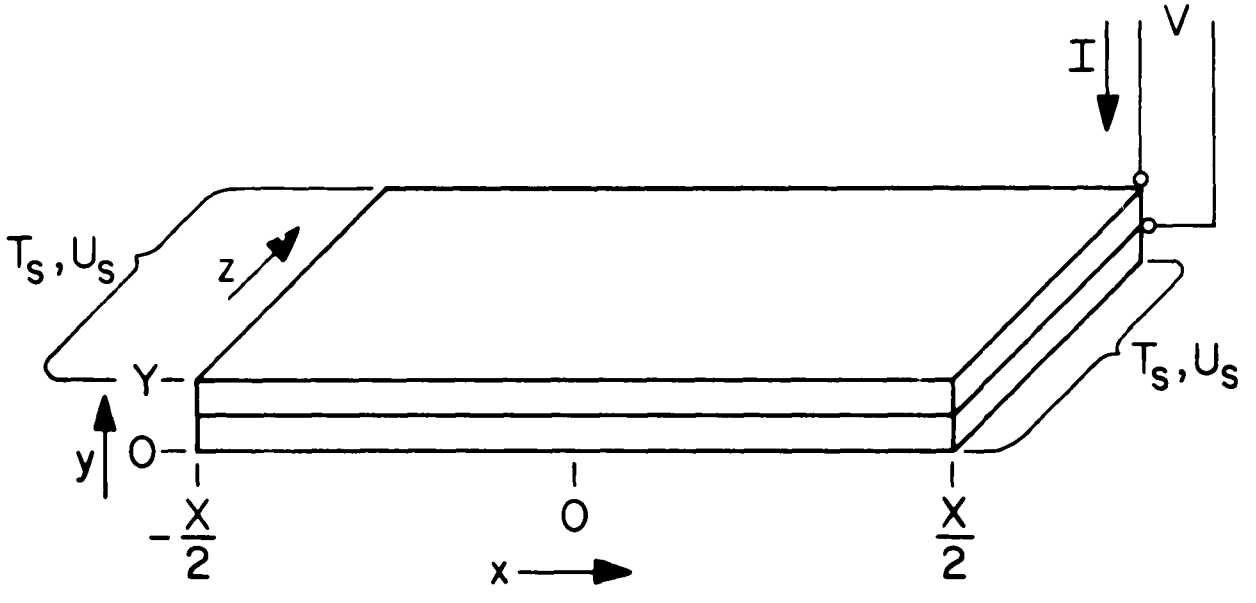


Figure 5.4

Structure with heat sinks at two ends
and heat flow along x-axis.

6. CURRENT AND TEMPERATURE DISTRIBUTIONS IN INTERDIGITATED STRUCTURES

6A. Introduction

In an interdigitated transistor structure the assumption of a uniform distribution of power density over working area is obviously incorrect. In the three parts of this section we shall consider the actual distribution of power. Subsection B takes into account the phenomenon of current crowding which causes the collector current to be concentrated near the outer edges of the emitter fingers. Subsection C considers temperature rises for this situation, and subsection D deals with an equivalent uniform power density.

Since in general in this section the equations will concern only the steady state case, corresponding to V_0, I_0, and P_0 of Section 2, and not a disturbed case like V_1, I_1 and P_1, the omission of the subscript "0" to simplify the equations causes no confusion. The subscript "0" is included only in references to previous sections where it is required.

6B. Current Crowding

In treating a transistor structure it may be necessary to take into account "current crowding" which results from the electric field in the base layer caused by the base current. In order to deal with this problem it is convenient to introduce not only currents per unit area but also currents per unit perimeter length where the perimeter concerned is that of the emitter-base junction at the edge of the emitter, represented by $x = 0$ in Fig. 6.1. For some cases of particular interest at high currents, the current from emitter to collector is effectively concentrated in a strip of width L near $x = 0$ and a negligible fraction flows deep under the emitter contact. The reason for the existence of an effective width L can be understood by a process of analysis involving defining appropriate currents and voltages and solving for the resulting distribution of current. The remainder of this subsection is devoted to this analysis.

Let the length of the emitter-base perimeter be L_{eb} and the total base current be I_b. Then in the base layer the current per cm flowing in the x-direction in Fig. 6.1 is B(x) where at the edge of the emitter we must have

$$B(0) = I_b / L_{eb} \tag{6.1}$$

The base current is usually chiefly consumed in the base layer by recombination with carriers from the emitter constituting the fraction

$(1 - \alpha)$ of the emitter current which does not reach the collector. Denoting the collector current density by $I_a(x)$ amp/cm^2 (so that the power dissipation is essentially $V I_a(x)$ where V is the emitter-collector voltage), the base current per unit area is $[(1 - \alpha)/\alpha] I_a(x)$.

We shall next derive two differential equations involving $I_a(x)$ and B(x) and their derivatives in respect to x. These equations also involve the resistance R_s in ohms per square of the base layer and α, the fraction of emitter current reaching the collector. Both α and R_s are assumed constant and independent of position x. (For extreme currents it would be necessary to include a dependence of α upon current.) The quantity U = kT/q of Eq. (2.2) is also introduced. The first differential equation expresses the fact that the base current must account for the emitter current that does not reach the collector, and the second differential equation expresses the fact that the emitter-base voltage changes with x and so consequently does $I_a(x)$, because of the electric field in the base layer needed to carry the current B(x). These two equations are then combined and solved to find the distribution of power $I_a(x) V$ over the transistor surface.

In the following equations, unless a specific value of x, such as x = 0, is involved, in general $I_a(x)$ and B(x) will be simply denoted by I_a and B, the dependence on x being understood.

The first differential equation is obtained by considering a strip

dx wide at x. In this strip the collector current per cm of perimeter is $I_a\,dx$. This requires a base current of $[(1-\alpha)/\alpha]\,I_a\,dx$. Consequently the change dB in B across this strip is

$$dB = (dB/dx)\,dx = -[(1-\alpha)/\alpha]\,I_a\,dx \tag{6.2}$$

This leads to one differential equation involving I_a and B:

$$I_a = -[\alpha/(1-\alpha)]\,dB/dx \tag{6.3}$$

A second differential equation involving I_a and B is obtained as follows: in keeping with Eq. (3.11), the logarithmic derivative of I_a in respect to emitter-base voltage is $1/U = q/kT$. The emitter-base voltage depends upon x because B(x) the base current per unit length flowing in the base layer sets up an electric field. This field is equal to R_s the resistance in ohms per square of the base layer times B(x), so that the change in forward base-emitter voltage in dx is

$$d[V_b(x) - V_e] = dV_b = -R_s B\,dx \tag{6.4}$$

where the emitter voltage is taken as independent of x because of the equipotential of the metal emitter contact. The change in voltage of Eq. (6.4) produces a change $d\,\ell n\,I_a$ given by

$$d\,\ell n\,I_a = -R_s B\,dx/U \tag{6.5}$$

or

$$d\,\ell n\,I_a/dx = -(R_s/U)B \tag{6.6}$$

This is the second differential equation mentioned above.

The two differential equations (6.6) and (6.3) can be combined into one equation for B by differentiating the logarithm of (6.3). As stated above, α is assumed independent of voltage and hence of x over the range of interest. The result can be written as

$$d^2B/dx^2 = -(R_s/U)B\,dB/dx \tag{6.7}$$

This differential equation can be directly integrated to give

$$dB/dx = (R_s/2U)(\text{const.} - B^2) \tag{6.8}$$

where "const." is a constant of integration. The solutions of Eq. (6.8) fall into three classes corresponding to positive, negative, and zero value for the constant of integration. For the value zero [the positive and negative cases are discussed below in connection with Fig. (6.4)] Eq. (6.8) integrates readily to give

$$1/B = (R_s/2U)(x + L) \tag{6.9}$$

where L is again a constant of integration which turns out to be the effective width for current crowding discussed above. In order to fit the condition at x = 0 that $B = B(x) = B(0) = I_b/L_{eb}$ of Eq. (6.1), the value of L must be

$$L = 2U/R_sB(0) \tag{6.10}$$

(At room temperature L is the distance across which the peak base-

layer field would produce a voltage drop of 50 millivolt.) In terms of L we may rewrite Eq. (6.8) as follows:

$$B(x) = (2U / R_s L) / [1 + (x/L)] = B(0) / [1 + (x/L)] \qquad (6.11)$$

This relationship is shown in Fig. 6.2; it is seen that the base current per cm flowing inwards under the emitter falls to half its value at x = L, to one third at x = 2L, etc.

The corresponding collector current distribution is obtained from Eq. (6.3):

$$I_a = [\alpha / (1 - \alpha)] B(0) / L[1 - (x/L)]^2 \qquad (6.12)$$

and the integral of this from 0 to x is

$$\int_0^x I_a \, dx = [\alpha / (1 - \alpha)] [B(0) - B(x)] \qquad (6.13)$$

The collector current density is shown in Fig. 6.3. The total current is the same as if a uniform current having the peak value

$$I_a(0) = \alpha B(0) / (1 - \alpha) L \qquad (6.14)$$

were to flow through a strip of width L. Actually half the collector current flows between x = 0 and x = L; between L and 2L, one-sixth of the current flows; between 2L and 3L, one twelfth, etc. Before considering the effects of this current distribution upon the power and the

temperature rise, we shall discuss other values of the "const." of Eq. (6.8).

For a negative value of "const." the B(x) vs. x curve always has a negative slope, as is shown in Fig. 6.4. Since the value of dB/dx depends on B^2, the curve is symmetrical about B = 0, which is shown as x = 4L for the example of Fig. 6.4. This situation can be considered to represent a symmetrical emitter finger of width 8L with base current per cm of B(0) flowing inwards, under the emitter (that is, in the +x direction) at x = 0 and again inwards (this time in the - x direction) at x = 8L with B(8L) = - B(0). For the case shown, dB/dx at x = 4L is about 1/3 of dB/dx at x = 0, so that the collector current drops by about a factor of three at the center of the emitter compared to the edge. This implies a value of about $-(1/2)B^2(0)$ for "const.". It also implies a case in which current crowding is not very severe. For severe current crowding, the "const." will be much smaller and the case of const = 0 will be an adequate approximation. (For intermediate cases Eq. (6.8) can be solved and the value of "const" chosen to fit the circumstances.)

For completeness, the case of const>0 will be considered. For this case, as x increases with dB/dx negative, B decreases and approaches exponentially a limiting value as shown. This corresponds to a situation, reminiscent of the junction transistor tetrode, in which current flows from one base contact to another base contact biased at a different voltage which produces a large reverse bias across the emitter junction.

The treatment presented here has omitted solutions of Eq. (6.7) which lead to dB/dx with positive values, such as would correspond to α values greater than unity in Eq. (6.3). These cases are not of the same general importance as those considered for heavy injection. Furthermore, in deriving Eq. (6.5) based on Eq. (3.11), only the current term involving the Boltzmann factor of forward emitter-base voltage, i.e., $\exp(V_b - V_e)/U$, was included. For heavy forward injection this approximation is usually valid unless the injected carrier density in the base becomes comparable to the majority carrier density. Modifications to include such effects can be made where their importance warrants the additional analytical complexity.

6 C. Current Crowding and Temperature Rises

In subsection 6.B it is shown that for severe current crowding the collector current density $I_a(x)$ is concentrated largely in a strip of width L. For purposes of analysis we introduce I_c the total collector current and I_p the collector current per cm of the L_{eb} cm of emitter-base perimeter; the subscript "p" signifies on a per cm of perimeter basis.

$$I_c = I_p L_{eb} = L_{eb} \int I_a(x)\,dx \tag{6.15}$$

The integral can be taken as extending from x = 0 to x = ∞ for severe crowding where L is much less than the emitter finger width. Otherwise it should extend half-way under the emitter. For the case of severe crowding, I_p can also be written as:

$$I_p = I_c / L_{eb} = \alpha B(0) / (1 - \alpha) \tag{6.16}$$

in keeping with Eq. (6.13) for the case of B(x) approaches zero.

The local power density $P_a(x)$ volts / cm^2 is

$$P_a(x) = I_a(x) V \tag{6.17}$$

where V is the emitter collector voltage. (The power produced by base current is generally negligible compared to $P_a(x)$.) The power density P_p on a per cm of perimeter basis is

$$P_p = V I_c / L_{eb} = V[\alpha / (1 - \alpha)]\, B(0) = V L I_a(0) \qquad (6.18)$$

in keeping with the above relationships between the current quantities.

The peak power density

$$V I_a(0) = P_p / L \qquad (6.19)$$

occurs only at x = 0. The total power is spread out as is $I_a(x)$ in Fig. 6.3. Thus the temperature rise will vary with x under the emitter finger. As an approximate average power density suitable for calculations, we replace the actual power distribution, which is spread out over many times L, by a restricted distribution, containing P_p watts per cm over a surface of extent $\pi L/2$ represented by a semicylindrical groove in Fig. 6.5. The average power density $(2/\pi)(P_p / L)$ is a compromise between the peak value P_p / L at x = 0 and the average value $(1/2) P_p / L$ over the strip $0 < x < L$. Over this semicylindrical surface, the uniform power density produces a substantially uniform temperature rise.

The thermal resistance from this semicylinder to a heat sink may be estimated from the construction shown on Fig. 6.5. For this purpose the edges of the emitter fingers are regarded as uniformly spaced with a period W; actually the spacings between strips L will alternate between values determined by the widths of the emitter fingers and the base fingers; W represents an average.

Thermal resistance for the semicylinder of diameter L to the heat sink Y below the surface may be readily estimated for the case where L << W. The problem of a set of line sources of thermal currents at uniform spacing may be dealt with by well known methods of potential theory using the complex transformation

$$w = \ln \sin z = \ln \sin (x + iy) \tag{6.20}$$

for which the temperature T corresponds to the real part of w and has limiting behaviors as follows:

$$\text{real part of } w = |y| - \ln 2 \qquad \text{for } |y| >> 1 \tag{6.21a}$$

$$\text{real part of } w = \ln r \qquad \text{for } r = |z| << 1 \tag{6.21b}$$

This leads to the conclusion that a current density leading to unit gradient at large values of $|y|$ has a potential difference between $|z| = r$ and $|z| = y_s$ of change in real part of

$$w = y_s - \ln 2 - \ln r \tag{6.22}$$

whereas for a uniformly distributed source producing the same unit gradient, the potential would be simply y_s; hence the ratio of thermal resistance, denoted by the symbol f_b, is

$$f_b = (y_s - \ln 2r) / y_s = 1 + (1 / y_s) \ln (1/2r) \tag{6.23}$$

In respect to Figure 6.5 it is evident that since the zeroes of sin z, corresponding to the axes of the semicylinders, are separated by

changes of x of π instead of W, we must have

$$y_s / \pi = Y / W \qquad \text{and} \qquad 2r / \pi = L / W \tag{6.24}$$

Hence the ratio of thermal resistance for the L, W, Y structure compared to uniform power on a layer of thickness Y is $f_b(L, W, Y)$ where

$$f_b(L, W, Y) = 1 + (1 / y_s) \ln (1/2r)$$

$$= 1 + (W / \pi Y) \ln (W / \pi L) \tag{6.25}$$

Since an average power density of $P_{a0} = P_p / W$ spread uniformly over the entire surface would produce, as discussed in connection with Eq. (5.3), a temperature rise of $P_p Y / Wc$, the same total power applied to the semicylinders at a rate P_p per cm of cylinder will produce a temperature rise $f_b(L, W, Y)$ larger. This will raise the stability index from a value s_a for uniformly distributed power by the same factor f_b, so that s_p the stability index for P_p per cylinder is

$$s_p = f_b s_a = a f_b(L, W, Y) P_p Y / Wc$$

$$= [1 + (W / \pi Y) \ln (W / \pi L)] a P_p Y / cW \tag{6.26}$$

In this expression we can interpret the fraction

$$cW / f_b Y = h_p \tag{6.27}$$

as the thermal conductance h_p per cm of periphery along the base-emitter junction so that s_p may also be written as

$$s_p = a P_p / h_p \tag{6.28}$$

where P_p / h_p is the temperature rise for quantities on a subscript "p" basis in U-units of temperature.

It is evident that the considerations of self-balance apply to this case so that if s_p is unity, a fluctuation of temperature over the entire surface at constant voltage will produce exactly enough extra power to sustain the needed thermal heat flow. However, if $s_p > 1$, excess power will be produced and instability will occur.

For severe emitter crowding L may be much less than W and $f_b(L, W, Y)$ may be greater than unity. The effect of this will be to cause the stability index at constant power level $I_p V$ to increase as V is decreased and I_p is increased, with a consequent decrease in L and increase in $f_b(L, W, Y)$.

6 D. The Distributed Interdigitated Structure

We next consider the problem of instability for the distributed interdigitated transistor structure. Since the temperature varies in the x-direction over the surface of Fig. 5.2, a simple u(x) variation is not suitable. It is appropriate, however, to introduce a formally identical function which represents the average rise in temperature over the surface. This average can be thought of as simply redistributing the power in the semicylinder of diameter L uniformly over a strip of width W to obtain a $P_{a0} = V I_p / W$. Because of the linearity of the heat flow equations and the principle of superposition, it follows that this average function u(x, y, z, t) satisfies the same boundary conditions at x = 0 and x = X and y = 0 discussed in equations (5.7) and (5.9), and of course the same diffusion equation (5.10); on the top surface y = Y where electrical energy is converted to heat requires a new boundary condition.

The effect of interdigitation and current crowding needs to be included in the y = Y boundary condition of Eq. (5.8) which relates the added power to the rise in temperature. The increase in temperature on the semicylinder will be larger than u(x, y, z, t) by the factor f_b of subsection 6.C. Hence the fractional increase in current density will be

$$i_a / I_a = a f_b(L, W, Y) u(x, y, z, t) \qquad (6.29)$$

This will produce an equal fractional increase in power density and hence in average power density P_{a0}. Consequently the boundary condition that should replace Eq. (5.7) becomes

$$c\,\partial u / \partial y = a f_b P_{a0} u(x, y, z, t) \qquad (6.30)$$

From this it is seen that the effect of interdigitation and current crowding is to affect the distributed structure treatment of Section 5 as if a were increased to a_p where

$$a_p = f_b a = [1 + (W / \pi Y)\, \ell n\, (W / \pi L)]\, a \qquad (6.31)$$

This is consistent with the result of Eq. (6.21) which shows also that a_p replaces a in the stability index.

This conclusion applies to disturbances that are uniformly spread over many fingers, as will be important for the most critical mode with the sin $(\pi x / X)$ factor and the frequency term ν_1. For much more localized disturbances approaches like those of Section 7 may be required.

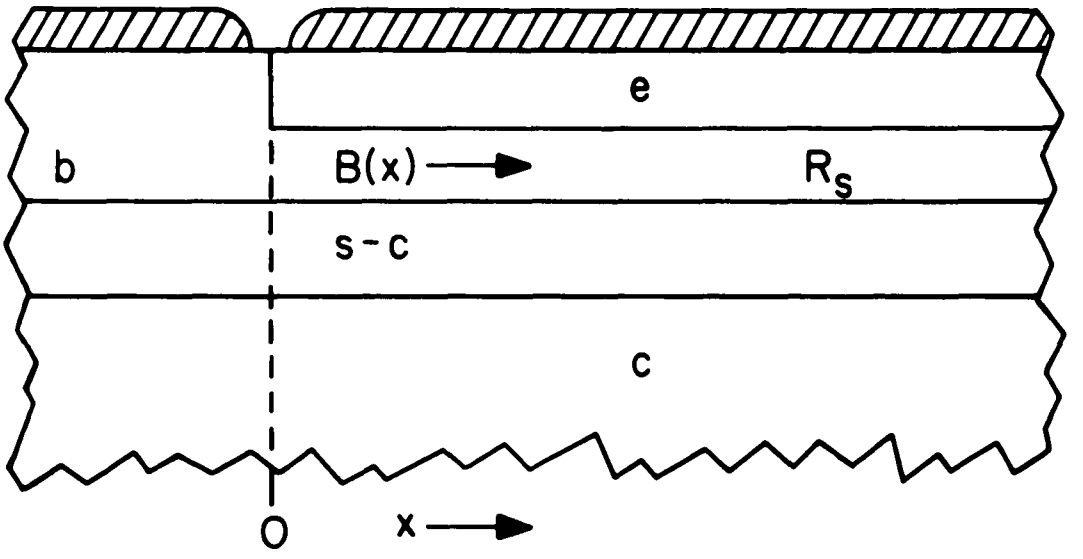


Figure 6.1

Model used to consider current crowding.

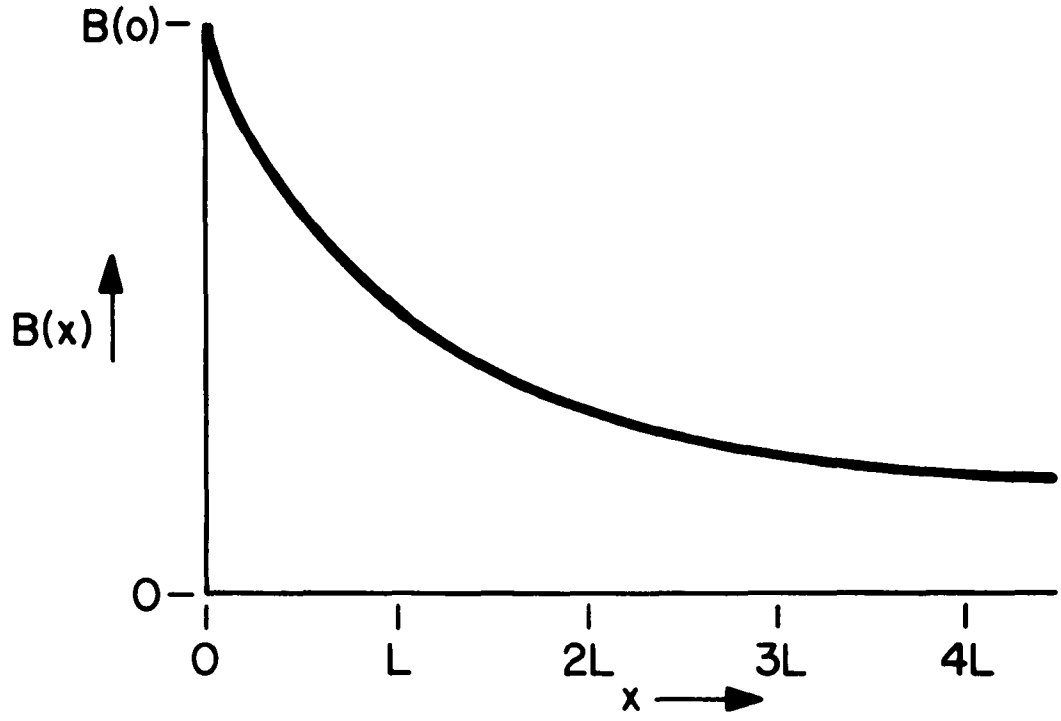


Figure 6.2

Dependence of base current upon distance under emitter.

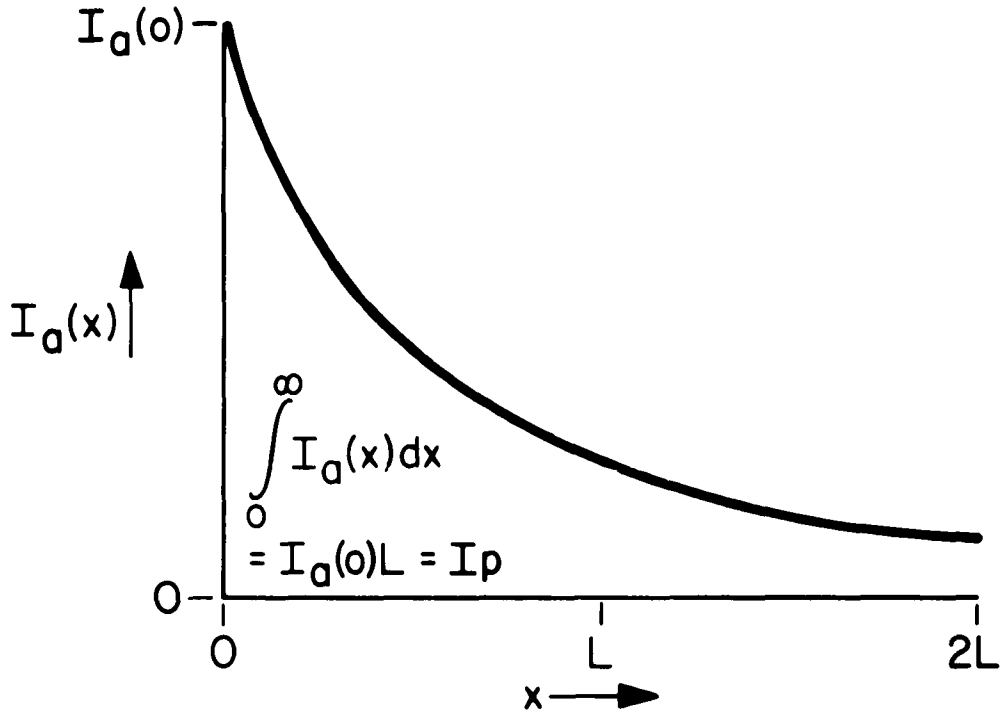


Figure 6.3

Distribution of current to collector.

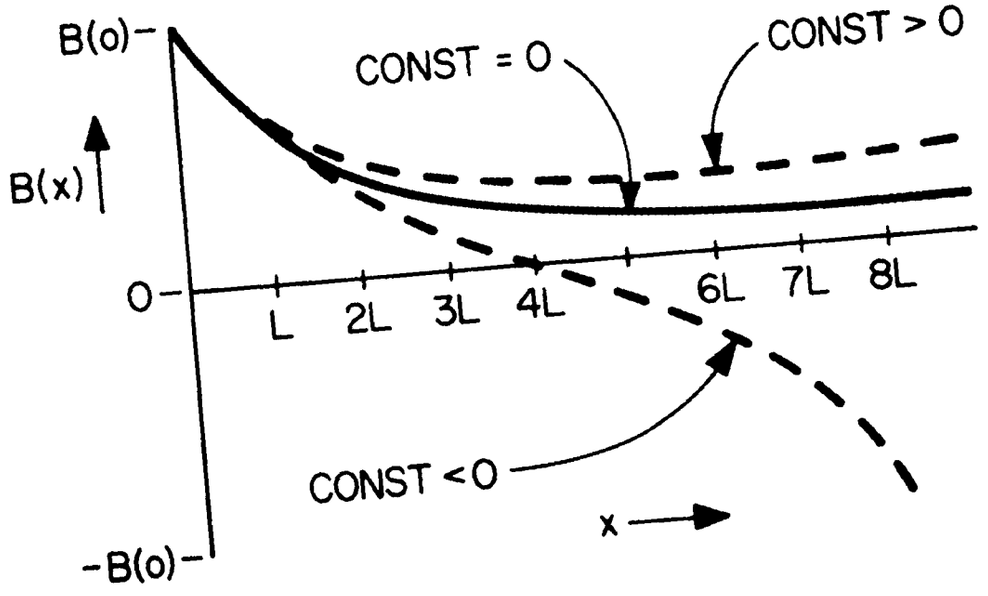


Figure 6.4

Variation of base layer current with distance for three values of "const."

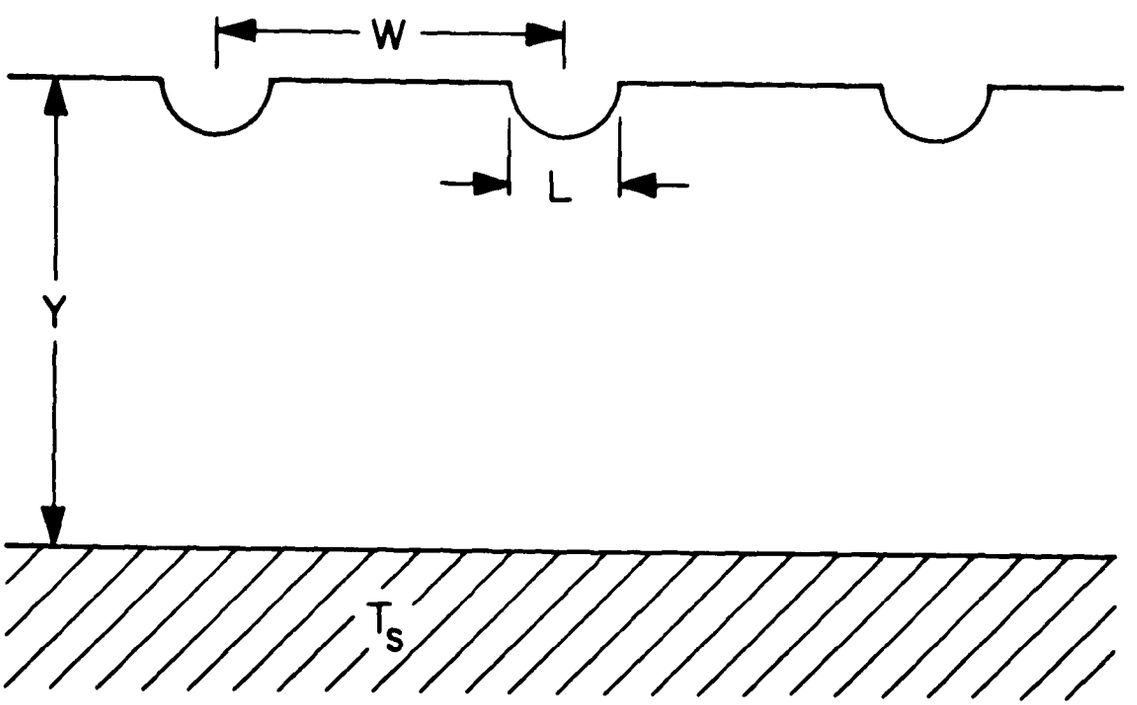


Figure 6.5

Construction used for estimating thermal conductance of interdigitated transistor structure.

7. LOCALIZED HOT SPOTS

An increase in temperature rise caused by a local concentration of electrical power can provoke thermal instability under circumstances in which a uniform distribution of power would not. Various defects in transistor structures may cause such local heating. For example, a thin spot in a base layer may cause a local concentration of collector current. Such thin spots may be caused by chemical contamination such as that causing diffusion pipes through a base layer or by diffusion down dislocations or groups of dislocations. Excessive reverse currents across the collector junction caused, for example, by defects in the space charge layer may also produce excess local temperature rises.

The effects of local heating are strongly dependent on quantitative aspects of the local disturbance, and no universal conclusions can be drawn save that in general thermal stability will be reduced. Some general aspects of the problem can be treated mathematically, however, so as to lay a foundation for treating particular cases.

In the treatment presented here it is assumed that the device may be divided into a local hot-spot region called A_1, having an area A_1, a local power density for the steady state condition of P_{a1} [we omit the subscript 0 for steady state introduced in Eq. (2.1), to avoid having three subscripts, such as $P_{a1\,0}$] and a heat conduction to the heat sink of h_1.

The total power applied to A_1 is $P_1 = A_1 P_{a1}$, and the disturbance in temperature from the steady state is u_1.

We shall assume that this small hot spot area is surrounded by a region called A_2, of area A_2, with local steady-state power density P_{a2} and heat conduction to the heat sink of h_2.

In order to deal simply with this problem, an artificial geometry is analyzed, illustrated in Fig. 7.1. Here a thin spot in the base layer illustrated in Part (a) is taken as equivalent to two concentric circular regions in Part (b), and these in turn are replaced by two concentric hemispheres, shown cut in half in Part (c). For this latter geometry uniform power generation on surfaces A_1 and on surface A_2 produces isothermal disturbances on A_1 and on A_2. This simplification serves the purpose of emphasizing the physical phenomena involved. A treatment based on more realistic geometry would lead to results that differed slightly quantitatively but would obscure the reasoning by great additional mathematical complexity.

If the power density in A_1 is $P_1 = P_{a1} A_1$ for the steady state, then at constant voltage and disturbance u_1 it will increase by

$$p_1 = P_1 a u_1 \tag{7.1}$$

where a is the temperature coefficient of current increase introduced in Eq. (2.3). The resulting temperature increase, caused by this dis-

turbance in A_1 alone, would be

$$u_1 = p_1 / h_1 \tag{7.2}$$

Eq. (7.2) will represent the condition of self-balance [discussed above following equations (2.13), (4.16) and (5.21)] if the power P_1 is so chosen as to make s_1 become unity:

$$s_1 \equiv P_1 a / h_1 = 1 \tag{7.3}$$

If the temperature rise due to P_2 is negligible, Eq. (7.3) becomes the stability condition.

The general stability condition when both P_1 and P_2 produce significant temperature rises can be formulated in terms of a self-balance condition like that for the two thermistors of Section 4.

For this purpose we shall use the concentric hemisphere geometry of Fig. 7.1. It is assumed that the heat sink is also concentric with radius x_s. For this case the disturbances p_1 and p_2 in heat flow due to temperature disturbances u_1 and u_2 are

$$p_1 = m(u_1 - u_2) \tag{7.4a}$$

$$p_2 = m(u_2 - u_1) + h_2 u_2 \tag{7.4b}$$

where heat flow theory leads readily to

$$m = 2\pi c x_1 x_2 / (x_2 - x_1) \tag{7.5}$$

$$h_2 = 2\pi c x_2 x_s / (x_s - x_2) \tag{7.6}$$

The heat conductance from A_1 to the heat sink is

$$h_1 = m h_2 / (m + h_2) \tag{7.7}$$

since for this geometry m and h_2 are in series between A_1 and the heat sink.

We shall first treat equations (7.4) for two limiting cases. The self-balance condition for power applied to A_2 only so that $p_1 = 0$ leads to $u_1 = u_2$ from (7.4a); and combining this with (7.4b), we obtain

$$p_2 = P_2 a u_2 = h_2 u_2 \tag{7.8}$$

$$s_2 \equiv P_2 a / u_2 = 1 \tag{7.9}$$

The self-balance condition for heat p_1 supplied to A_1 and $p_2 = 0$ is also readily found from equations (7.4) to lead to

$$p_1 = P_1 a u_1 = m h_2 u_1 / (m + h_2) = h_1 u_1 \tag{7.10}$$

$$s_1 \equiv P_1 a / h_1 = 1 \tag{7.11}$$

These results, which include duplicating Eq. (7.3), are simply a check of the correctness of equations (7.4) for the limiting cases in which only one or the other of A_1 and A_2 has a power source.

If powers P_1 and P_2 are both applied, the self-balance condition requires that equations (7.4) be simultaneously satisfied when rewritten in terms of u_1 and u_2 by expressing p_1 and p_2 as $p_1 = P_1 a u_1 = s_1 h_1 u_1$ and $p_2 = s_2 h_2 u_2$:

$$(s_1 h_1 - m) u_1 + m u_2 = 0 \tag{7.12a}$$

$$m u_1 + (s_2 h_2 - m - h_2) u_2 = 0 \tag{7.12b}$$

These equations can be satisfied if and only if the determinant of the coefficients of u_1 and u_2 vanishes. This algebraic result can be manipulated to be put in the form

$$s_1 = (1 - s_2) / [1 - (h_1 / m) s_2] \tag{7.13}$$

This form of the self-balance condition lends itself to interpreting the mutual influence of the interacting regions A_1 and A_2 as is discussed below.

We note first that in Eq. (7.13) we may use Eq. (7.7) to rewrite h_1/m:

$$h_1 / m = h_2 / (m + h_2) < 1 \tag{7.14}$$

so that the denominator of Eq. (7.13) does not vanish as s_2 approaches unity. Hence, as s_2 approaches unity, the right side of Eq. (7.13) approaches zero; this leads correctly to the self-balance condition $s_1 = 0$, $s_2 = 1$ for heat applied to A_2 only, as discussed with Eq. (7.9). Similarly, Eq. (7.13) is satisfied by $s_1 = 1$, $s_2 = 0$ in keeping with Eq. (7.10).

For any given value of s_2 lying in the stability range $0 < s_2 < 1$, the smallest value of s_1 satisfying (7.13) and producing self-balance is obtained if h_1/m approaches zero so as to make the denominator of Eq. (7.13) as large as possible; now h_1/m can approach zero if x_1 approaches x_2. [See Eq. (7.5)] For this condition Eq. (7.13) leads to

$$s_1 + s_2 = 1 \tag{7.15}$$

The physical interpretation of $h_1/m = 0$ is simply that A_1 and A_2 coalesce into one and the same surface to which a power $P_1 + P_2$ is applied, leading to an effective s value for that surface of $s_1 + s_2$.

For all other cases with $h_1/m > 0$, it follows that self-balance requires

$$s_1 > 1 - s_2 \qquad \text{or} \qquad s_1 + s_2 > 1 \tag{7.16}$$

As a specific example, we consider a relatively small hot spot so that $20x_1 = 2x_2 = x_s$. For this case

$$h_2/m = x_2 x_s (x_2 - x_1) / (x_s - x_2) x_1 x_2$$

$$= 10 \times 20 \times 9 \;/\; 10 \times 1 \times 10 = 18 \tag{7.17}$$

$$h_1/m = 0.95 \tag{7.18}$$

For this case, if the large area A_2 has $s_2 = 0.5$, then the hot spot area must have

$$s_1 = 0.5 / (1 - 0.475) = 0.95 \tag{7.19}$$

Hence local instability will build up at area A_1 at much the same power level as if A_2 had zero power applied.

For the example considered the ratio in current densities P_{a1} / P_{a2} required to have $s_1 / s_2 = 0.95 / 0.5 = 1.9$ is obtained by noting that since the same voltage V_0 is applied to A_1 and A_2, the ratio of power densities is I_{a1} / I_{a2} so that

$$1.9 = \frac{s_1}{s_2} = \frac{I_{a1} A_1 h_2}{I_{a2} A_2 h_1} = \frac{I_{a1} \times 1 \times 18}{I_{a2} \times 100 \times 0.95} = \frac{0.19 I_{a1}}{I_{a2}} \tag{7.20}$$

or

$$I_{a1} = 10 I_{a2} \tag{7.21}$$

The total current to area A_1 is, however, only 0.1 that to area A_2. The influence of the hot spot in increasing instability can then be described as follows: if in a uniform device, 1% of the area is made to carry an increase of 10% in total current, the local power produced will cause instability even though the area A_2 by itself is below instability by a factor of 2.

It is evident that methods like those of Section 5 can be applied to analyze correctly special cases with geometries like those of Fig. 7.1(b). Although the results will be analytically accurate, they will not change the general conclusion that a local hot spot may

significantly reduce stability if s_1 for this spot per se approaches unity at lower power levels than does s_2 for the general area by itself. For large ratios of size between A_1 and A_2, a close approach to unity is required either by s_1 or by s_2 to produce instability. Also, for large ratios of areas, much higher power densities in the small area will be required to cause it to provoke instability; however, these larger localized densities will make only a small contribution to total current and total power.

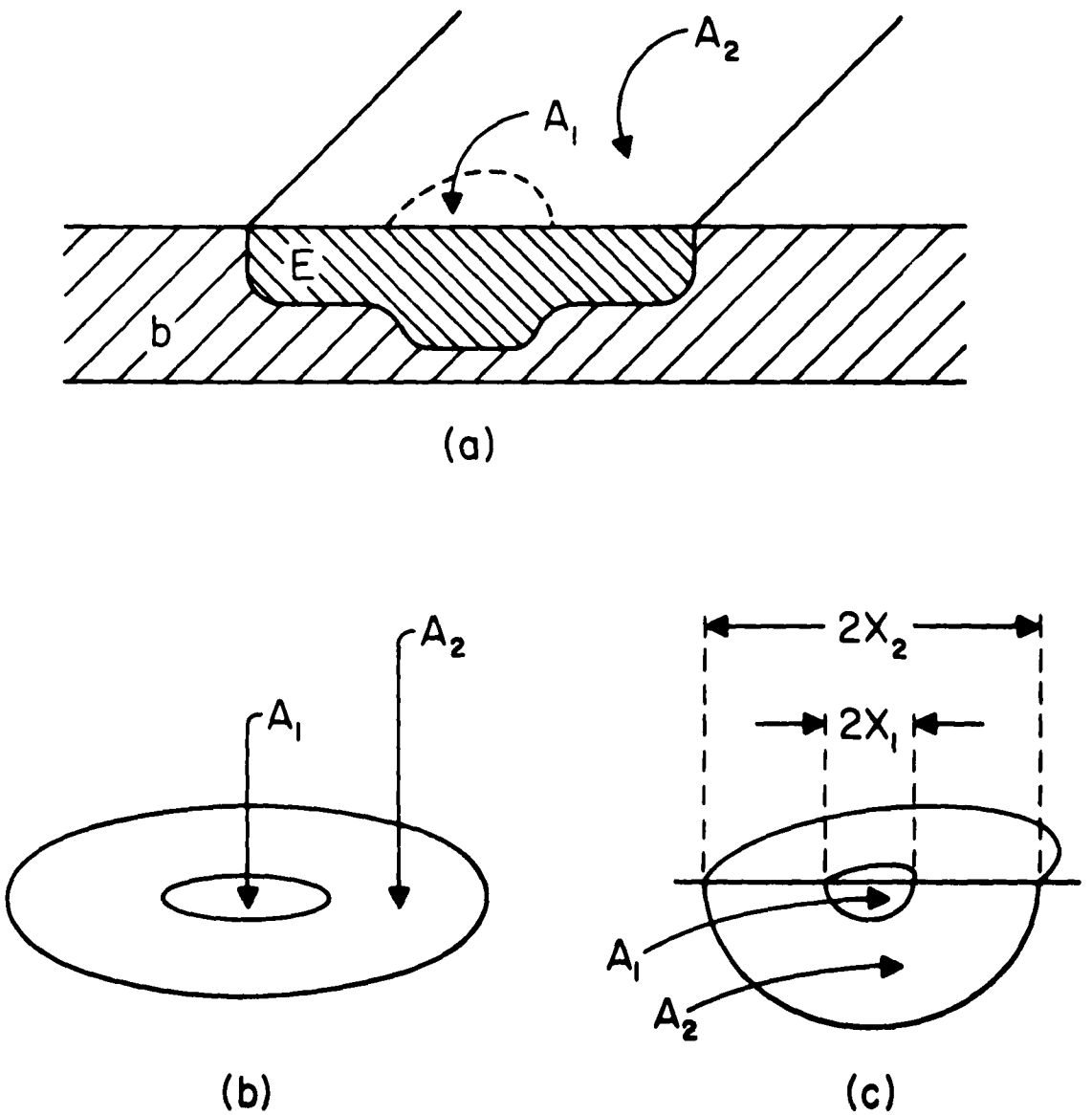


Figure 7.1

Effect of a localized heat generating region.

Experimental Results

8. EVIDENCE FOR LOCALIZED HOT SPOTS

All experiments to be described were performed on a developmental npn silicon power transistor. Fig. 8-1 shows this transistor. Its size is 1.4mm by 2.7mm. The npn structure is made by diffusion into epitaxial material. The evaporated aluminum emitter and base contacts are interdigitated. The emitter, for example, consists of a central "backbone" and 2 x 16 fingers extending from it. Four thermocompression bonds are made to the backbone to serve as emitter contacts. One such contact is made to the base. The package used for this developmental transistor is a TO-3, which provides an excellent heat sink.

This transistor is intended to carry maximum currents of the order of 5 amperes. The base layer is about 1.5μ thick for good high-frequency response. The thermal resistance from junction to case, measured under conditions where the heat generation is uniform over the active area, is about $2^{\circ}C$ per watt.

Evidence for the formation of localized hot spots was first found while measuring the thermal resistance of the transistor. Fig. 8-2 gives the circuit used for these measurements.

In this circuit a pulse of power is applied of sufficient length for the transistor to reach thermal equilibrium. The pulse is then cut off and the emitter-base junction voltage V_{EB} is observed while a small measuring current is flowing. This junction voltage has a temperature coefficient at constant current of the order of 2.3 mv per degree C, and can hence be used as a

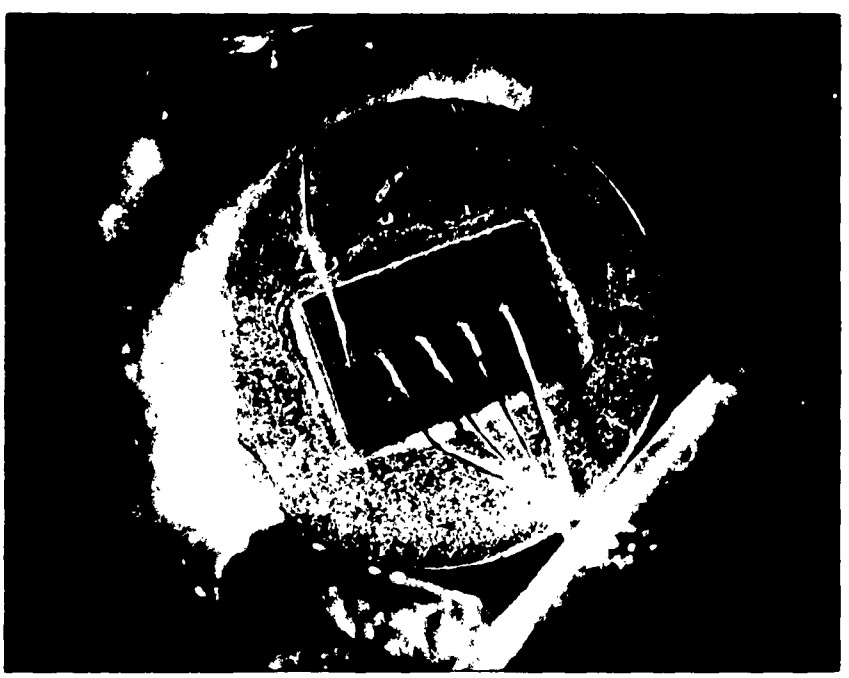
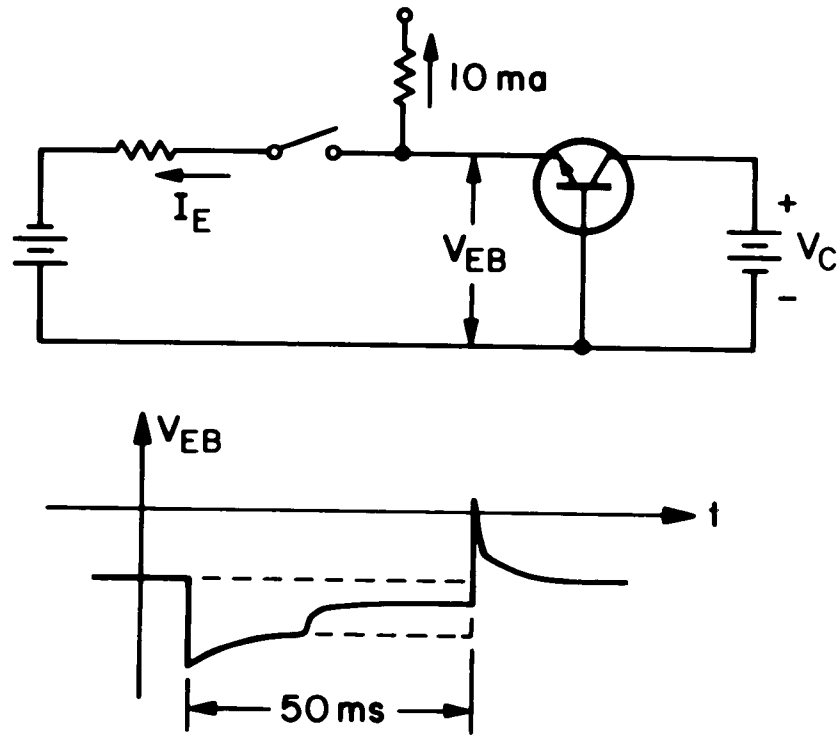


Figure 8.1

Developmental silicon power transistor used for instability studies.

thermometer to measure internal temperature rise. This experimental value is in good agreement with theory. When the power pulse is applied by means of a large emitter current pulse with constant collector voltage, $|V_{EB}|$ increases owing to the higher forward drop required at the larger current. As the device heats up, $|V_{EB}|$ decreases until thermal equilibrium is reached, because a smaller junction voltage is required to sustain a given current at higher temperature. The thermal time constant of the device under observation is of the order of 10 ms.

When a critical power level was reached, a step in V_{EB} occurred as shown in Fig. 8-2. This step first occurs a few thermal time constants after the main pulse is applied, and indicates that a small portion of the device must be getting hot and carrying most of the current, because the reduction in voltage is sufficient to bias off the remainder of the emitter junction. The subsequent spike which appears when power is cut off is partly due to the decay of thermally generated current from the hot region flowing through the base sheet resistance to the base contact. This current causes a lateral voltage drop along the base layer, with the interior portion of the base near the hot spot region being positive with respect to the base contact which is at ground potential. Since the emitter is forward biased at this time, but carrying a small current, its potential cannot be more than 0.5 volts or so below the most positive portion of the base layer. In effect, the emitter potential follows that of the interior base region, and one observes from emitter to ground nearly the entire lateral drop of the base layer. This voltage decays very rapidly as the hot spot cools, since the thermally generated current decreases exponentially with temperature.



EMITTER-BASE VOLTAGE
DURING HOT SPOT FORMATION

Figure 8.2

Experimental circuit used for measuring temperature rises and hot spot development

By simultaneously observing the voltage spike at the end of the pulse and the reverse base current at the same instant, one determines the effective spreading resistance through the base layer from the hot region. A typical value for the operating points reported here is 1000 ohms. In an experiment on a special structure performed in connection with development of this transistor, the base sheet resistance was measured to be 3000 ohms per square at room temperature. A spreading resistance of 1000 ohms can arise in this layer if the current is generated within an area whose radius is about 1/4 of the distance to the base contact. The diameter of such a region is in the order of 30μ. This value is in reasonable agreement with the 100μ diameter estimated by temperature-sensitive paint (reported later), considering that most of the current will originate toward the center of the region, and that the base sheet resistance may be considerably higher at the high local temperatures involved.

There is an interesting mode in which failure can occur due to the above mechanism in a circuit like that of Fig. 8-2. As the power is raised, the hot spot becomes hotter, the thermally generated current increases rapidly, and a higher lateral base voltage drop results. The emitter-base junction is barely forward biased at the hot region, so that this junction is reverse biased in the region around the base contact by approximately the amount of the lateral drop. When this lateral drop reaches the reverse emitter-base breakdown voltage (about 8 volts in the present transistor) avalanche current can flow at the edge of the junction. This current is carried by injection at the hot region, and tends to maintain the temperature. In effect,

a path for emitter current is provided to ground, and the resulting very low impedance from collector to base will draw excessive current and destroy the transistor unless current in the collector circuit is limited. Confirmation of this model is given by the fact that when this mode of conduction is taking place, light emission from large numbers of microplasmas is observed along the entire periphery of the emitter-base junction.

For one particular example to be discussed, the step first appears at an emitter current of 600 milliamperes and 34 volts collector voltage (see Fig. 8-2). The time it takes to appear when it is first observed is of the order of a few thermal time constants. At higher voltage and thus higher power, it occurs much sooner. For example, at 38 volts it occurred in a few milliseconds and at 40 volts the device was destroyed, showing an emitter to collector short.

Observations were made of the voltage at which a hot spot began to develop for various currents. A typical locus of voltage-current points is shown in Fig. 8-3. It is noted that the locus from about 25 to 50 volts roughly corresponds to a constant power level, but at lower voltages considerably higher powers can be tolerated before thermal instability occurs.

The electrical measurements just described yield only an indirect proof for the existence of a spot with enhanced temperature. However, with a temperature indicating paint a "hot spot" can be seen to develop when the step in emitter base voltage appears. A photograph of a hot spot developed on heat sensitive paint on one transistor is shown in Fig. 8-4. The center region of this hot spot is about 100μ in diameter, and from the characteristics of the paint it is estimated that the local temperature is of the order of 300 or 400°C.

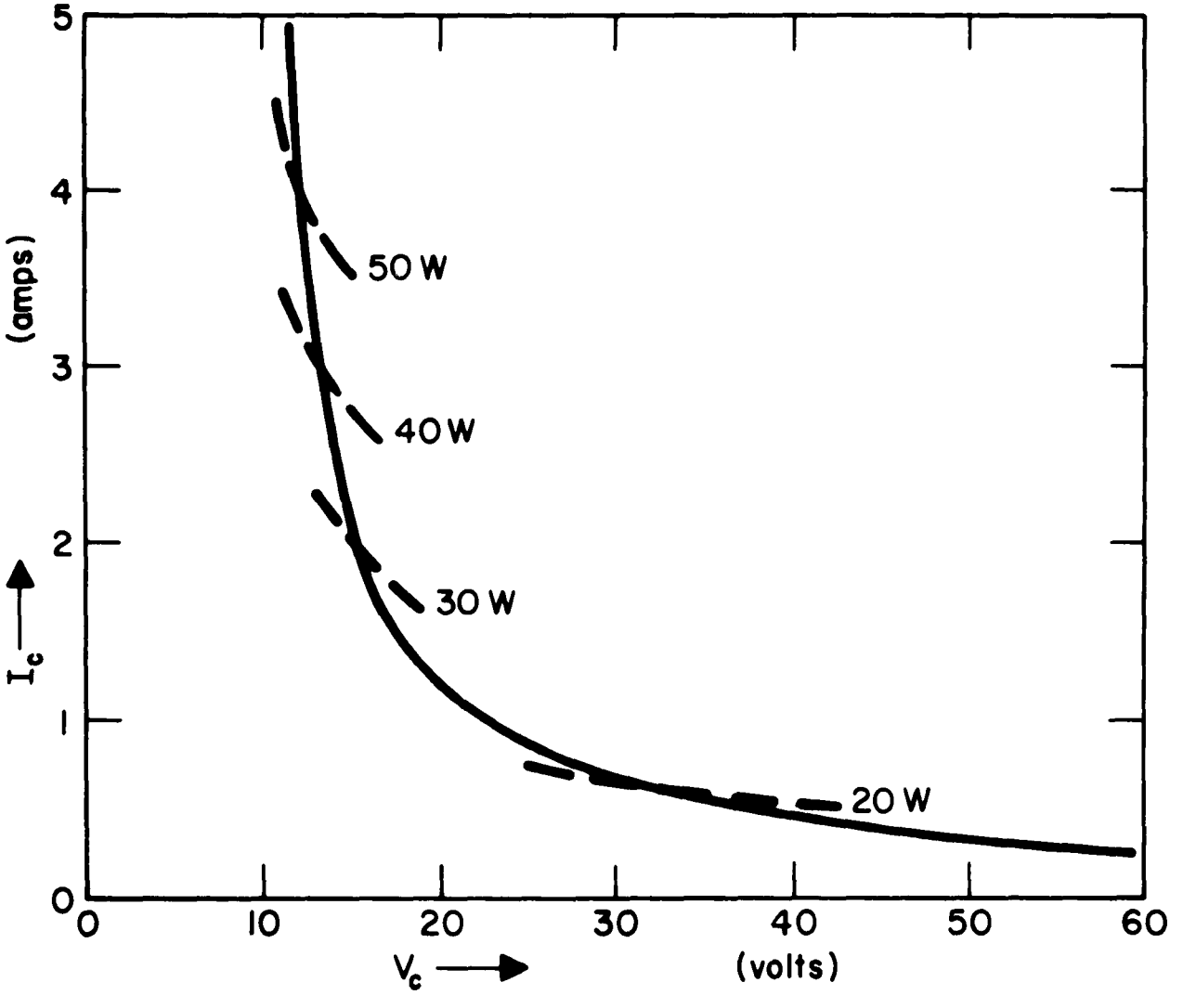


Figure 8.3

Locus of voltages and currents for the onset of "hot spot" formation in a power transistor

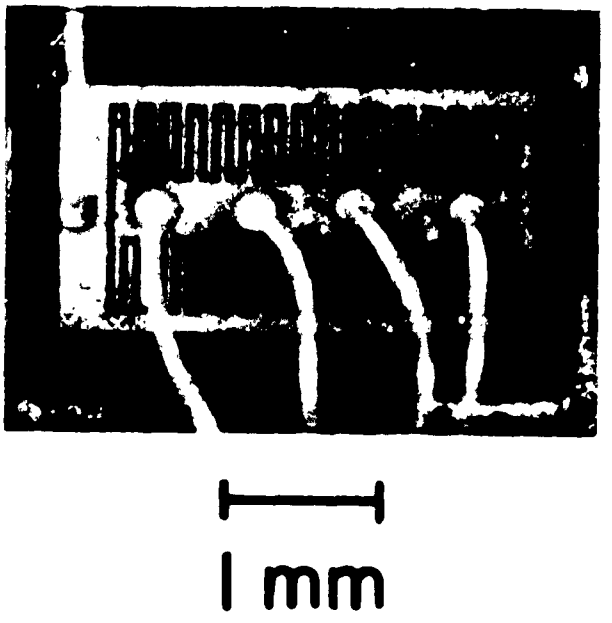


Figure 8.4

Power transistor structure covered with temperature sensitive paint. The location of the hot spot is seen by the discoloration below and to the left of the second emitter lead from the left

At the total power being dissipated, one would expect only a 40°C temperature rise if the power were distributed uniformly across the device. The observed diameter of the order of 100μ is consistent with the model that the hot spot size is determined by the spreading resistance of the collector body region. The collector body on this transistor is a 20μ thick layer of 5Ω-cm material. It is believed that when the hot spot develops almost all of the collector voltage appears across the collector body as an ohmic drop, and the size of the hot spot is such that the spreading resistance from the hot spot region into the collector body is equal to this ohmic drop divided by the emitter current.

Several hundred transistors were tested for the occurrence of the hot spot. All transistors showed the electrical effect described by Fig. 8-2. The spread in power level necessary to create the "spike" was approximately ±10% for a given constant collector voltage V_c. This critical power level depends strongly upon collector voltage as indicated in Fig. 8-3. The hot spot develops more easily at low currents than at high currents. It is believed that this is because of a stabilizing influence of the resistance of the collector body region which limits local power generation at the lower voltages and higher currents.

If the power level is increased beyond the onset of the hot spot the transistor will be permanently damaged and eventually fail. In most cases the failure mode is a short circuit between emitter and collector. Several of such destroyed transistors were closely examined. Highly localized damaged areas were almost always found. The transistors showed pits or grooves of

molten silicon. This clearly indicates a localized area of very high temperature, in accordance with our model. A typical picture of such a failure is given in Fig. 8-5. A molten channel can be seen to connect emitter and base. The observations are similar to those described and cited by French and Schafft[1] for damage by "second breakdown."

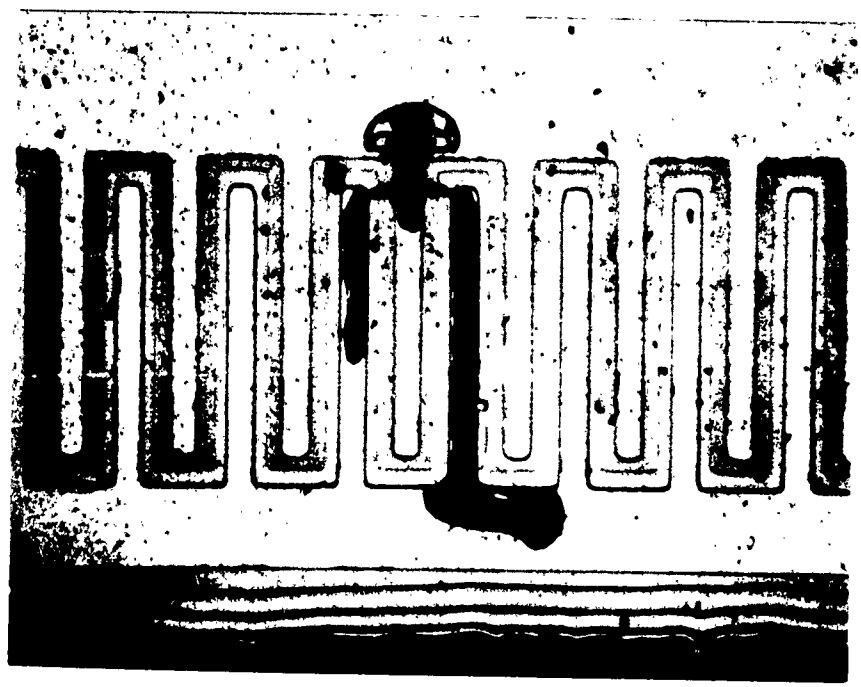


Figure 8.5

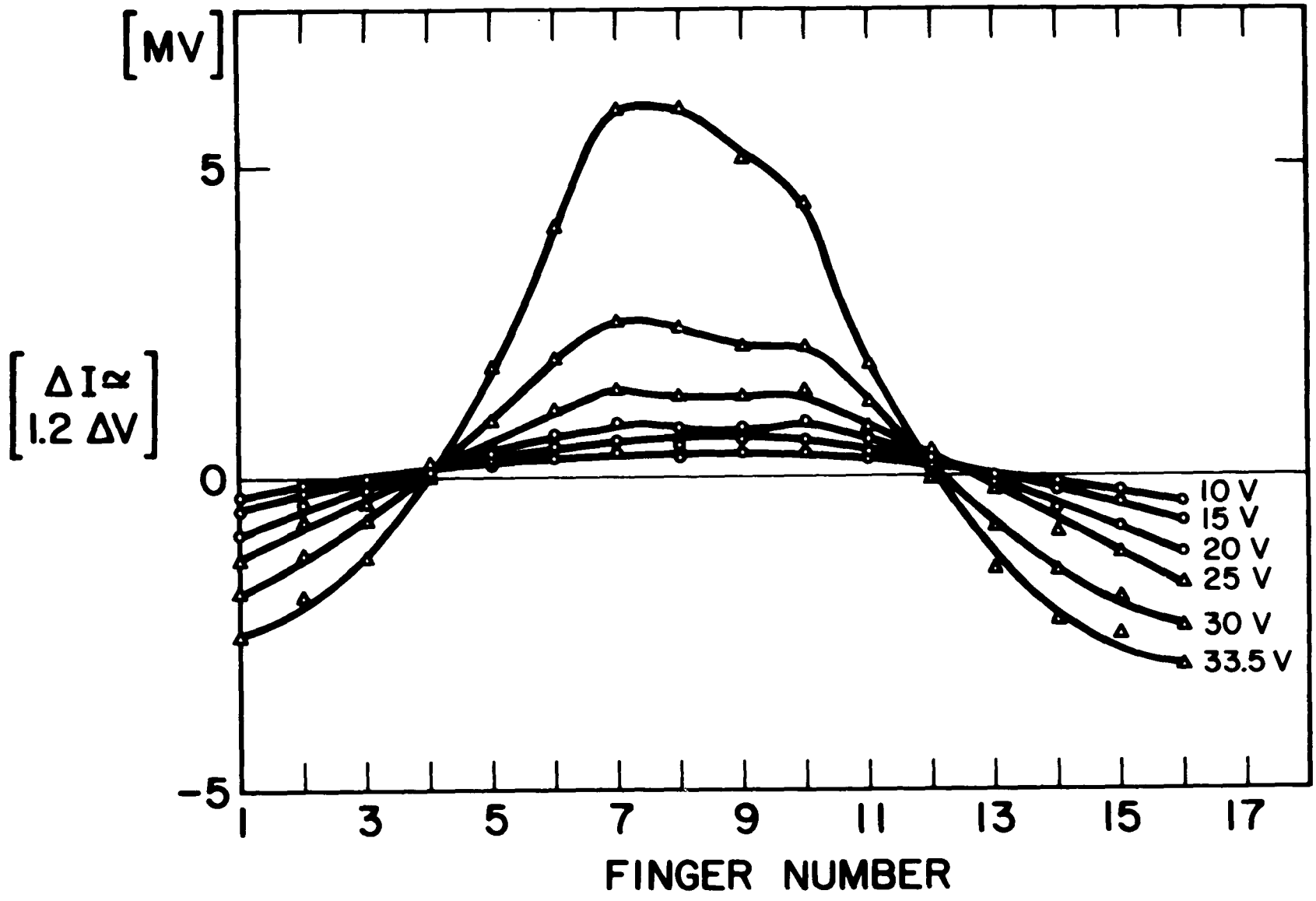
Portion of a transistor permanently damaged by excessive power. A channel of molten material is seen where the silicon had melted.

9. CURRENT DISTRIBUTION MEASUREMENTS

By potential probing of the evaporated aluminum emitter contact the current distribution across the transistor can be measured. Probes are placed at the base and tip of each finger of the structure to indicate the current being drawn by each finger times the average length to which it flows in the finger. We will first show the static distribution across the entire transistor prior to hot spot development. Then we will show how current builds up with time in two individual fingers.

Fig. 9-1 shows the change in finger voltages from their uniform initial values for the 16 fingers on one side of the backbone across the entire length of the device. The emitter current is 600 milliamperes and distributions are shown for various collector voltages, all of these producing less power than that needed to form a hot spot. The initial finger voltages are about 6 millivolts so that finger No. 8, for example, is approximately doubling its voltage when the collector voltage is 33.5 volts. It is seen that the distribution looks like the $n = 3$ cosine mode discussed in the first chapters. (see Eq. 5.11) It is believed that these distributions represent initial unstable build-up current which is brought to a halt by the non-linearities in the device. When the power level is pushed sufficiently high, in this case 600 milliamperes times 34 volts, the instability builds up to such an extent that most of the current ends up flowing through two adjacent fingers.

The way in which current varies with time can be seen in Fig. 9-2. This shows photographs of finger voltage wave forms. One is for one of the two fingers where the hot spot ultimately develops, and the other is for a



DISTRIBUTION OF CURRENT IN EMITTER FINGERS

Figure 9.1

Change in finger voltage during a pulse for each of the sixteen emitter fingers on one side of the backbone of a transistor like Fig. 8.1 before the voltage step of Fig.8.4 has developed. The finger voltage is measured as a voltage drop between two probes, one at the base of the finger at the backbone and the other near the finger tip. The total current was 600 ma and various collector voltages are shown as parameters for the curves.

7-5-62
50283

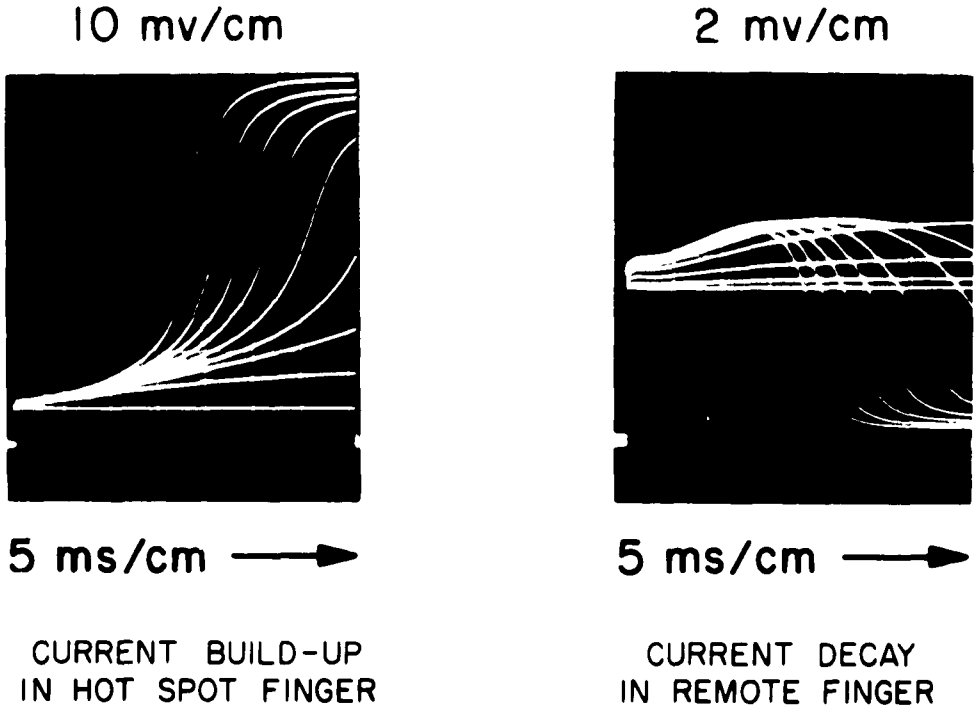


Figure 9.2

The collector voltages for the nine traces for voltage drop on one of the two hot spot fingers are 0, 30, 33, 33.5, 34, 34.5, 35, 35.5, and 36 volts. From the trace prior to the pulse application it is seen that the voltage drop increases nearly ten times; this is sufficient to account for practically all the current flowing on the two hot spot fingers to a region near the backbone. For the remote finger, four fingers distant from the hot spot fingers, the eleven traces correspond to collector voltages of 0, 10, 20, 30, 33, 33.5, 34, 34.5, 35, 35.5, and 36 volts. It is seen that for V_C = 36 volts the initial drop of about 7 mv decreases by more than a factor of 10 at the same time as the drop in the hot spot finger increases from about 8 mv to 80 mv.

finger which is four fingers distant. At low voltages corresponding to conditions in Fig. 9-1,initial build-up of current is apparent in both fingers. When the hot spot forms, the current increases to a limiting value in one finger and decreases practically to zero in all the others. This occurs at the same time as the step in the emitter base voltage waveform occurs in Fig. 8-2. The limiting value of current which can be carried by a finger depends on the collector voltage and the resistance of the collector body region. At low collector voltage, the maximum finger current is correspondingly small. A large current is then forced to spread over many fingers and a hot spot is prevented from forming.

So far we have considered the effects of lateral thermal instability when the emitter current is constrained to remain constant by the external circuit. In a situation where the base current is controlled rather than the emitter current, as in a common emitter circuit, we believe that when thermal instability has proceeded to a point where a hot spot is developed, the phenomenon of secondary breakdown is observed. The intense local heat of the hot spot provides a localized thermal generation current into the base layer which stimulates further emission from the emitter in that region. The thermally generated current may be estimated as follows:

For example, if the hot spot is taken as having an area of the order of a(hundred microns)^2 and a thickness of forty microns, giving a volume V of about 4×10^{-7} cm^3, at 400°C, the intrinsic density is $n_i = 3 \times 10^{16}$ cm^{-3} and if the lifetime at that temperature is $\tau = 10^{-7}$ sec, then the thermally

generated current is estimated as $I = qn_i V/\tau = 20$ ma, which is enough to sustain collector currents of the order of an ampere. Temperature rises of at least 300 to 400°C are quite reasonable on the basis of the thermal resistance of silicon, which varies almost inversely with temperature,[4] and the size of the hot spot region. Furthermore, Tauc's observation[5] that local temperature rises to 600°C to 780°C do not damage p-n junctions suggests that higher rises than 780°C do occur when the transistor is damaged.

Thermally generated currents in reverse biased p-n junctions associated with very high temperature rises have been reported by J. Tauc.[5] Tauc observes hot spots at temperatures of 600°C to 780°C associated with "Thermal Breakdowns in Silicon p-n Junctions." Tauc does not consider thermal generation rates. However, for these temperatures the intrinsic densities are between 10 and 100 times larger than those estimated above for 400°C. Hence, they can easily carry the necessary currents. Tauc also observed a V-I characteristic which is nearly hyperbolic corresponding to constant power input. This is expected for a thermistor-like action with high activation energy. Similar hyperbolic V-I curves are frequently observed in secondary breakdown, and this fact lends further support to the thermal generation model for secondary breakdown. Studies of thermal generation rates at these high temperatures have apparently not been carried out and will be needed for quantitative analysis of the effects involved.

---

[4] R. G. Morris and J. C. Hulst, Phys. Rev. 124, 1426 (1961).

[5] J. Tauc, Phys. Rev. 108, 936 (1957).

In any event, for thermally generated currents of the order of 20 ma or more, as estimated above, no external base current is then required to maintain emission, and a localized thermal runaway occurs. The current concentrates, and the voltage falls to a value where the local power generation is just sufficient to cause the thermally generated current required to sustain emission. By means of temperature indicating paint, exactly the same sort of hot spots as that shown in Fig. 8-4 have been observed when the transistor is allowed to enter the second breakdown condition rather than developing the hot spot as in Fig. 8-2. Furthermore, the critical points in the collector current-voltage characteristic required to trigger secondary breakdown essentially coincide with the current and voltage values required to develop a hot spot in a circuit like Fig. 8-2. A typical locus of critical points is like that shown in Fig. 8-3.

10. FACTORS INFLUENCING HOT SPOT FORMATION

Several series of experiments were performed to determine what electrical, thermal and structural parameters of the transistor can be correlated with the hot spot effect.

No correlation was found between the power required to produce a hot spot and any of the conventional electrical parameters such as the junction breakdown voltages, collector to emitter sustaining voltage, current gain (beta) or saturation voltage. As is to be expected, there is correlation with thermal resistance (junction to case) since a unit with high thermal resistance will require less power to reach a given temperature. Furthermore, high thermal resistance is generally caused by poor alloying of the device to its mounting base, and such alloying is usually non-uniform over the active device area. Thus, portions of the device which are over areas of incomplete alloying will have a high local thermal resistance, and these portions will evidently be favored in the formation of hot spots. Attempts to deliberately introduce high local thermal resistance are reported below.

The arrangement of the four emitter leads exerts an influence on the position of the hot spot formation. In most cases the hot spot developed in the center region of the structure, close to the backbone and also close to one of the emitter contact leads. Apparently the series resistance of the emitter structure is of importance. By using only two emitter contact leads, one at each end of the backbone, approximately 10% more power could be handled, indicating that this arrangement of leads resulted in a more uniform

temperature distribution. This was confirmed by measurements like those of Fig. 9-1.

Further effects of the emitter contact resistance were demonstrated by cutting one or more of the four emitter leads. If the lead nearest the hot spot location is cut, the hot spot will generally move to be near one of the intact leads. This process can be repeated until only one lead is left. The interpretation of this changing hot spot location is that the voltage drop along the emitter sheet resistance tends to bias-off portions of the transistor remote from an emitter contact lead, particularly at high currents, so that conduction and temperature rise is greatest in the vicinity of an emitter lead. If more than one lead is present, the hot spot will appear near the one which has a "weak spot", evidence for the existence of which is reported below.

The geometry of the transistor structure is influential for the formation of hot spots as is to be expected from the model. A smaller transistor structure with only 6 fingers instead of 32 gave a power density handling capacity of about twice that of the large structure before a hot spot was created. The smaller area of this transistor could be cooled more effectively and more uniformly. The reason for this more effective cooling is that the lateral dimensions of the small device are comparable with the distance through which heat must flow to an effective heat sink (about 300-400μ). The mutual thermal conductance between two regions of the device is thus larger compared to the thermal conductance to the heat sink than in the large device. Thus, greater lateral stability is expected in accordance with the discussion of two

thermistors in parallel in Section II-4 (see Figs. 4-1 and 4-2). Similar conclusions are reached in Section II-5 on the distributed model in connection with Figs. 5-1, 5-2 and 5-3.

An "inverted" transistor structure was investigated. This structure had the base contact in the center and the emitter on the periphery. Otherwise the structure was identical to the one shown in Fig. 8-1. This arrangement should result in somewhat better cooling of the emitter area. The units tested showed an increase of about 10% for the power level where hot spot formation sets in.

Devices that had high thermal resistances between the transistor and its metal base showed onset of hot spots at lower power levels than the average of transistors with good contacts to the heat sink. This evidence indicates that thermal effects are prevailing -as hypothesized- and that cooling of the transistor must be made as effective as possible in order to attain higher power levels without hot spot formation. Deliberate increase of the thermal resistance was done by drilling holes into the molybdenum disks on which the transistors were mounted. These units gave hot spots consistently at drastically reduced power levels compared to the average of the regular devices. On the other hand, it was found that hot spot formation can be retarded by placing a wire probe directly on top of the region where the hot spot develops. It is believed that the wire probe serves as an additional heat sink.

There is experimental evidence for so-called "weak spots" in the transistor structure, where the current localization is most likely to be initiated.

Localized external heating was provided with a hot tantalum wire on the transistor while it was under test with the circuit of Fig. 8-2. It was found that there were regions where the hot spot could be introduced more easily than in other regions. The sensitive regions were in general those where the hot spot formed of its own accord in the absence of external heating.

On some transistors it was observed that the location of the hot spot would spontaneously change between pulses (see Fig. 8-2) after one location had been hot for perhaps half a minute or more. The new location was generally near another of the emitter contact leads. On one unit, the hot spot jumped back and forth between two locations. It is speculated that there may be a number of "weak spots", and when one of them becomes hot, it can be temporarily or permanently made less sensitive. The hot spot will then move to another weak region. Since temperatures considerably higher than $400^{\circ}C$ are thought to exist at a hot spot, it is evident that some alteration of the region with time can be expected.

The nature of these sensitive regions could not yet be determined. It may be speculated that these areas have defects such as non-uniform doping, thinner base layers, precipitates, or dislocations. Definitive experiments about such defects have been initiated but could not be completed under the present contract.

11. CONCLUSIONS

The theoretical and experimental work described indicates that a lateral thermal instability is present in large area transistors and presents a serious limitation for performance and reliability of these devices. This limitation is particularly severe for class A applications where a steady operating point of large power dissipation is required. However, the fact that the hot spot requires collector voltages of the order of 30 volts (for the particular transistor under investigation) and times of order of tens of milliseconds to develop has the practical significance that hot spots do not develop when the transistor is used for switching or class C applications. For these, the operating point passes so rapidly through the unstable region that the hot spot does not have time to form. Thus, a power transistor can dissipate and control much larger powers if the operation is essentially a switching one rather than a class A operation in which both power dissipation and relatively high collector voltages are simultaneously required.

To conclude, we observe that lateral thermal instability can be a serious limitation on the power handling capacity of transistors, particularly in thin, broad-area, high-frequency power transistors. It seems likely that the secondary breakdown phenomenon observed in nearly all transistors is triggered by the sort of basic thermal instability described in this paper. It is believed that the secondary breakdown condition is maintained by thermal generation of base current in the heated region of the collector junction.

Appendix

DISLOCATIONS AND SEMICONDUCTOR DEVICE FAILURE*

H. J. Queisser

Shockley Transistor, Unit of Clevite Transistor

Palo Alto, California

Abstract

Dislocations cannot be avoided in conventional semiconductor devices. Even if dislocation-free starting material is used, there will be dislocations introduced through the solid state diffusion steps which are necessary to form layers of different conductivity types. The influences of the dislocations upon device properties and failures are discussed in this paper.

Direct electrical effects of dislocations are surprisingly weak. No detectable changes in p-n junction characteristics are found in diodes having high dislocation density of a small-angle grain boundary. These results are inconsistent with the assumption of "dangling bonds" at the dislocation.

A strong influence is exerted by dislocations upon the distribution of impurities. Diffusion is enhanced along dislocations. This can lead to serious doping inhomogeneties in devices, resulting in failures such as emitter-collector shorts or "weak

* This work supported by Contract AF 30(602)-2556, Rome Air Development Center.

spots" in base layers.

Precipitation of foreign metals occurs preferentially at dislocations. These precipitates cause undesirable "softness" of p-n junctions. Oxygen impurity atmospheres have been found to affect the electrical behavior of dislocations in silicon. Dislocations should not be regarded separately from their surrounding impurity atmospheres; their combined effects influence the properties and reliability of semiconductor devices.

I. INTRODUCTION

In this talk we shall consider dislocations in crystals and their influence upon performance and reliability of semiconductor devices.

I would like to restrict myself to silicon devices, such as transistors, diodes, or solar cells. Silicon is an elemental semiconductor with a comparatively simple crystalline structure. It can be made very clean and highly perfect. It has been extensively investigated as a semiconductor, at the same time it is an almost ideal material for investigations of lattice defects, including dislocations. We therefore suggested for this "Physics of Failure Program" to study the complex question of dislocation-induced failures only in silicon. There is hope that the results found with this substance can be generalized to also cover phenomena in more complicated materials such as compound semiconductors or piezoelectric materials.

We proposed to study some basic physical phenomena involving dislocations in silicon, such as diffusion, precipitation, or photoeffects. This approach seems preferable and should precede empirical-statistical methods applied to series of actual devices. From our results on these basic phenomena we can make predictions about device failures, which might be induced by dislocations; about these things I would like to talk today.

Before we enter this discussion, however, I would like to speak

very briefly about dislocations in general. Next, I want to give you an introduction to the specific problems that arise in silicon, which crystallizes in the diamond structure. Methods to generate and observe dislocations are particularly interesting for the homopolar semiconductors; we shall briefly cover this subject. It is possible today to obtain silicon crystals without any dislocations. It seems to be obvious that for highly reliable devices one should work exclusively with this perfect silicon. Unfortunately, this material does not improve device parameters, because in making a device, dislocations will be introduced almost by necessity. I will show that by solid state diffusion, one can generate considerable numbers of dislocations in the crystalline material forming the device. Other handling and processing steps act in a similar way. Thus, in practice, one cannot avoid these lattice defects, and the questions arise: what harmful effects do the dislocations have; how can these effects be reduced or avoided; and finally, how can dislocations be utilized to our advantage. I can only talk about the first two aspects. We shall see that the direct electrical effects of dislocations are surprisingly weak. Some of our recent results lead to speculations why this is so. Indirect effects, however, can be of importance for device failure. Diffusion enhancement and precipitation are especially dangerous, which I hope to show in some more detail after the brief introduction to dislocations.

II. DISLOCATIONS IN THE DIAMOND LATTICE

A dislocation is a one-dimensional lattice defect. Along the dislocation line, the lattice is out of register, such as indicated in Fig. 1. This figure shows a so-called edge dislocation in a primitive-cubic lattice. Inserted into an initially perfect crystal is an extra half-plane of atoms. This half-plane terminates at the dislocation. A region of compression is generated above the "core" of the dislocation, and a region of dilatation of the lattice below the "core". The symbol $\perp$, used in Fig. 1 denotes such an edge dislocation.

Silicon has the more complicated diamond crystal structure. The dislocations in this lattice are also more complex. Fig. 2 shows a model of a dislocation in silicon. A row of terminating spokes is seen in the center. These symbolize broken and unsaturated Si-Si bonds. If each dislocation required such "dangling" bonds, strong electrical effects must be expected[1]. However, this electrical disturbance may be avoided by a diffusional rearrangement[2], such as shown in Fig. 3 for another type of dislocation. Rings of five and of seven atoms each are formed here; all bonds are saturated. It is not yet clear which of these two distinctly different types of dislocations is present in silicon. We shall see that there is reason to doubt that the simple picture of this row of dangling bonds is correct.

Dislocations can easily be detected in silicon[3]. The most convenient method is by etching. Each dislocation can be made visible by a pit on the surface. Other methods of detection are given by the

electron microscope, by x-ray methods and by infrared transmission or birefringence. These techniques are standard procedures for quality control checks in silicon device production. Today's crystal growing techniques are so refined that one can get absolutely dislocation-free material[4].

It may sound surprising that device production is not based upon such highly perfect material. Empirical tests indicate that this dislocation-free material does not yield transistors or diodes superior to those of silicon with moderate dislocation densities. This problem was analyzed a few years ago by a panel of experts at an AIME meeting[5]. Dislocations in semiconductors were found to be surprisingly harmless; their effects seemed to be much smaller than those of as yet uncontrolled production techniques.

In the meantime, more information has been obtained. Some results were also found under this "Physics of Failure" program, which throw some light on these puzzling problems. One discovery was that during the fabrication of the device, dislocations can be introduced in numbers which greatly exceed the initial concentrations [6,7]. Diffusion of undersized dopant atoms, such as phosphorus or boron, gives rise to dislocation formation. Fig. 4 shows the surface of a boron-diffused slice after etching. A network of lines is observed. These lines run along crystallographic [110] directions. They are slip lines, indicating that the lattice was so severely stressed by the diffusing boron that slip was initiated. Slip means that a dislocation moves into the lattice.

X-ray scanning[8] of such boron-doped surfaces has revealed these dislocations (see Fig. 5). High concentrations of undersized substitutional dopants are needed to yield these "diffusion-induced" dislocations. But such high concentrations are very commonly used in producing transistors, diodes, or solar cells. We thus have one effect which could explain why dislocation-free material did not yield any superior devices, simply because the material was no longer free of dislocations after the device was made. Even if the diffusion leaves the lattice intact, other steps such as alloying, thermocompression bonding, or scribing will cause damage and dislocations. It must be concluded that in most practical cases dislocations will be present in the device. We must therefore find out how the dislocations affect a device.

III. DIRECT EFFECTS OF DISLOCATIONS

The assumption of "dangling" bonds at dislocations in the diamond structure has one immediate consequence. These unsaturated bonds should act like acceptor atoms, trying to capture an electron. The dislocations should provide extra conductivity. This would be undesirable in a reverse-biased p-n junction, where one wants to reduce the reverse current to as low a value as possible. The dislocations with dangling bonds should definitely increase reverse currents in diodes. This has not been observed, however. Diodes were made with a controlled high number of dislocations running perpendicular to the p-n junction. This was achieved

by using a crystal with a small angle grain boundary. Such a boundary consists of a regular array of evenly spaced parallel edge dislocations, in our case, about 5 lattice parameters apart. These diodes, as depicted schematically in Fig. 6, do not show any anomalies. Their behavior is no different from diodes on the same slice, but off the boundary. This absence of strong conduction effects speaks against the simple model of a dislocation with dangling bonds.

Other direct electrical effects of dislocations include influence on carrier lifetimes, and recombination and breakdown phenomena. It has been established that dislocations reduce the lifetime of minority carriers. It is not possible, however, to give an unambiguous numerical description of this effect, since it depends not only on the dislocation itself, but also upon the impurities surrounding the dislocation. In the next section we shall discuss this important interplay between dislocations and impurities in more detail.

IV. INDIRECT EFFECTS OF DISLOCATIONS

Edge dislocations attract foreign atoms and surround themselves with an "atmosphere" of these impurities. The attraction of impurities toward dislocations is plausible if one considers that an oversized impurity fits much better into the dilatation region of the dislocation. A smaller impurity is easier to accomodate in the compression region than in the undisturbed lattice. This attraction mechanism leads to effects which

can damage the electrical performance of semiconductors. Two of these effects shall be discussed here: the enhancement of dopant diffusion, and the precipitation of impurities.

Foreign atoms, including donors and acceptors, diffuse faster along edge dislocations than in undisturbed material[9]. We believe that there are three reasons for this. At first there is the attraction mentioned above, which leads to an increase in the concentration of the diffusant. Secondly, there may be a similar enhancement of vacant lattice sites around the dislocation. This would enhance any diffusion with a vacancy-mechanism. Finally, the activation energy for the jump process may be lowered at the dislocations because of the altered environment. These explanations have been offered to describe the diffusion along the dislocations of small angle grain boundaries in silicon. Fig. 7 shows a typical result of a diffusion along such a boundary. A phosphorus diffusion creates an n-type layer at the surface. This can be made visible by beveling, and staining with an acid solution. The original p-type material is stained dark while the diffused layer appears light. It is seen that at the grain boundary, the n-p junction is driven in much further than in the undisturbed material. A spike-shaped diffusion front results.

Three layer structures at grain boundaries give exaggerated examples of device failures which may be caused by the diffusion enhancement along the dislocation. Fig. 8 gives a "base layer short". The phosphorous "emitter-diffusion" broke through the diffused p-layer,

creating a short circuit between the emitter and collector regions. This is possible because different dopants have different diffusion enhancements, even at an identical diffusion temperature. The size of the diffusing atom is of prime importance for the magnitude of this enhancement.

Gallium diffuses even more rapidly along dislocations than phosphorous. Fig. 9 shows that a gallium diffusion penetrated through the entire thickness of a 100 micron slice at the boundary, while the penetration depth in the regular material is only one-tenth this value.

Fig. 10 gives another example of highly non-uniform doping around the dislocations of a small-angle grain boundary. Gallium, an acceptor atom, was diffused into a silicon slice. Subsequently the sample was heat treated without a gallium source. Out diffusion occurs under these conditions. The original n-type behavior is reinstalled close to the surface because the gallium evaporates out of the silicon lattice. At the grain boundary, this process is greatly enhanced, caused by the rapid diffusion to the surface.

Isolated dislocations should, in principle, give similar effects, however much less pronounced than seen by the superimposed effects of the dislocations at the grain boundary. But even slight non-uniformities in doping may create rather serious failures in devices. High frequency transistors have base layers thinner than 1 micron. It is clear that such a structure is very sensitive to a non-uniformly penetrating diffusion front. A thin spot can be built up in such a transistor. This may lead

to a concentration of current and a localized temperature rise. Such localized "hot spots" severely limit the operation of transistors. This problem will be treated in a later paper by R. Scarlett and W. Shockley.

Aside from the doping non-uniformity which arises from dislocations, there is another serious consequence of the impurity atmospheres. Contaminants, especially heavy metals, are frequently present in semiconductor devices. If these metals are in solid solution they may be comparatively harmless. If, however, metal precipitates are formed in p-n junctions, a drastic failure results. The reverse currents of these junctions are raised, often to such a degree as to make the device unusable. This phenomenon is called "softness". It has been shown by Goetzberger and Shockley[10], that softness can indeed be correlated with metal precipitates. Precipitation requires nucleation centers. Dislocations provide such centers. Therefore, softness is more probable for higher dislocation densities in a device. This statement has been proved by experiments on grain boundary[11] diodes and also on twin boundaries[12]. Second-order twin boundaries in silicon provide favorable nucleation sites for precipitates; diodes with such boundaries were found to be predominantly soft, while diodes on good material within the same slice showed the desirable " hard" reverse current-voltage characteristic.

This precipitation mechanism can lead to a gradual destruction of an initially good device. Operating temperatures are generally well below those required for an appreciable diffusion of donors or acceptors.

However, fast diffusing interstitial metals can migrate freely enough, especially along dislocations, to form the harmful precipitates during the lifetime of the device. Estimates for the times and temperatures involved must be based on precise numerical data about such diffusions. These numbers have not yet been determined; the basic experiments are more complicated than those for donor and acceptor diffusions.

Finally, I would like to mention one more precipitation phenomenon, which may seem more elusive, but which probably is of importance for the effects of dislocations in silicon. We have done some investigations which strongly suggest that oxygen "atmospheres" around dislocations are decisive for electrical effects[13]. If a grain boundary is illuminated with a small light spot, a photo-voltage can be observed, similar to the photoresponse of a p-n junction solar cell. In silicon it was found that this photo-voltage depends drastically upon the thermal history of the sample.

Changes in polarity and magnitude of the photoresponse indicate that the electrical character of the boundary dislocations varies. These variations closely parallel effects found previously[14] in oxygen-containing bulk silicon. Oxygen is present in crucible-grown silicon in quantities up to 10^{18} cm^{-3}. These oxygen atoms can be dissolved; they can form SiO_4^+-complexes, or SiO_2 precipitates. The kinetics of the various transitions between these formsare well known[14]. These mechanisms could explain the varying photoresponse. We thus hypothesize that

oxygen surrounds the dislocations and determines what the effects of the dislocations are. All these "indirect" effects of the dislocations indicate that one should regard the dislocation, not as an isolated entity, but always in combination with its impurity atmospheres.

V. CONCLUSIONS

We conclude from the evidence just presented that dislocations seem hardly avoidable for practical silicon devices. Dislocations cause harmful and undesirable effects and endanger the reliability of a device. Failure will result mostly from indirect effects, especially through doping non-uniformities caused by diffusion enhancement and precipitation of metals at dislocations which can destroy p-n junctions. Strong direct effects should be expected if the "dangling" bond hypothesis were correct. Such effects are missing. This lack may be explained by two assumptions. Either there are no broken bonds at all; the dislocations have structures such as shown in Fig. 3. Or, secondly, the dangling bonds are saturated by impurities, especially oxygen. In this case, dislocations and impurities would mutually annihilate in part their undesirable electrical effects.

I have attempted to present a short review of the problems which dislocations cause for device reliability. Many aspects could not be touched upon. For example, how do the recently discovered stacking faults[15] in epitaxial silicon affect a device? We have some ideas that they should

not act too differently from dislocations, since we have a fairly good physical picture of these faults from which we can deduce what must be important for device failure. Many questions are still left unsolved, but we should have a fair chance to obtain answers, especially in a program such as our common one which stresses the need for basic research.

References

1. W. Shockley, "Dislocations and Edge States in the Diamond Crystal Structure", Phys. Rev. 91, 228 (1953)

2. J. Hornstra, "Dislocations in the Diamond Lattice", J. Phys. Chem. Solids 5, 129 (1958)

3. For an extensive recent review, see V. Sadagspan, "A Review of Plastic Deformation of Semiconductors", Solid State Abstracts 3, 90 (1962)

4. W. C. Dash, "Growth of Silicon Crystals Free from Dislocations", J. Appl. Phys. 30, 459 (1959)

5. W. E. Taylor, W. C. Dash, L. E. Miller and C. W. Mueller, "The Role of Dislocations in Device Properties" in: "Properties of Elemental and Compound Semiconductors", edited by H. C. Gatos, Interscience Publishers, New York and London, 1960 p. 327

6. H. J. Queisser, "Slip Patterns on Boron-Doped Silicon Surfaces", J. Appl. Phys. 32, 1776 (1961)

7. S. Prussin, "Generation and Distribution of Dislocations by Solute Diffusion", J. Appl. Phys. 32, 1876 (1961)

8. G. H. Schwuttke and H. J. Queisser, "X-Ray Observations of Diffusion-Induced Dislocations in Silicon", J. Appl. Phys. 33, 1540 (1962)

9. H. J. Queisser, K. Hubner and W. Shockley, "Diffusion along Small-Angle Grain Boundaries in Silicon", Phys. Rev. 123, 1245 (1961)

10. A. Goetzberger and W. Shockley, "Metal Precipitates in Silicon p-n Junctions", J. Appl. Phys. 31, 1821 (1960)

11. H. J. Queisser in Final Report for Contract AF(33 (616)-7786, "Solar Cell Parameter Study", (1962)

12. H. J. Queisser, "Properties of Twin Boundaries in Silicon", submitted to J. Elecrochem. Soc. (1962)

References (cont'd)

13. W. W. Hooper and H. J. Queisser, "Photoresponse of Small-Angle Grain Boundaries in Silicon", Bull. Am. Phys. Soc. 7, 211 (1962)

14. W. Kaiser, H. L. Frisch, and H. Reiss, "Mechanism of the Formation of Donor States in Heat-Treated Silicon", Phys. Rev. 112, 1546 (1958)

15. H. J. Queisser and R. H. Finch, "Stacking Faults in Epitaxial Silicon", J. Appl. Phys. 33, 1536 (1962)

Figure Legends

Fig. 1 Edge dislocation in a primitive-cubic lattice .

Fig. 2 Dislocation in the diamond lattice, showing "dangling" bonds.

Fig. 3 Edge dislocation in the diamond lattice without "dangling" bonds.

Fig. 4 Evidence for diffusion-induced dislocations by slip lines, which appear after etching of a heavily boron-doped [110]-silicon surface.

Fig. 5 X-ray topogram showing diffusion-induced dislocations.

Fig. 6 Mesa diode at a small-angle grain boundary. The dislocations of the boundary run perpendicular to the p-n junction.

Fig. 7 Diffusion enhancement by the dislocations of a small-angle grain boundary. Beveled and staining section shows spike-shaped profile of the diffusion front.

Fig. 8 Model of a shorted base layer of a transistor, obtained by two diffusions on a small-angle grain boundary.

Fig. 9 Penetration of a gallium diffusion through a 100 micron thick silicon slice at the grain boundary. Regular junction depth is only 10 microns.

Fig. 10 Preferential out-diffusion of Gallium at the small-angle grain boundary.

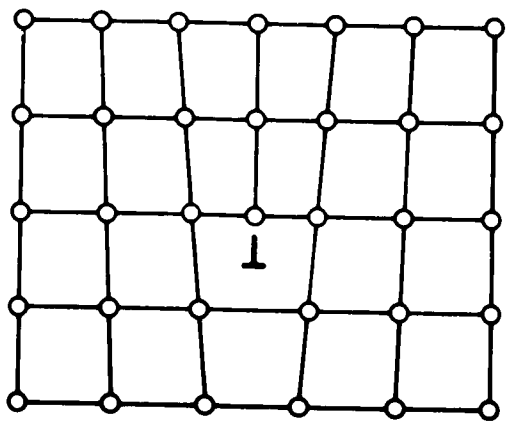
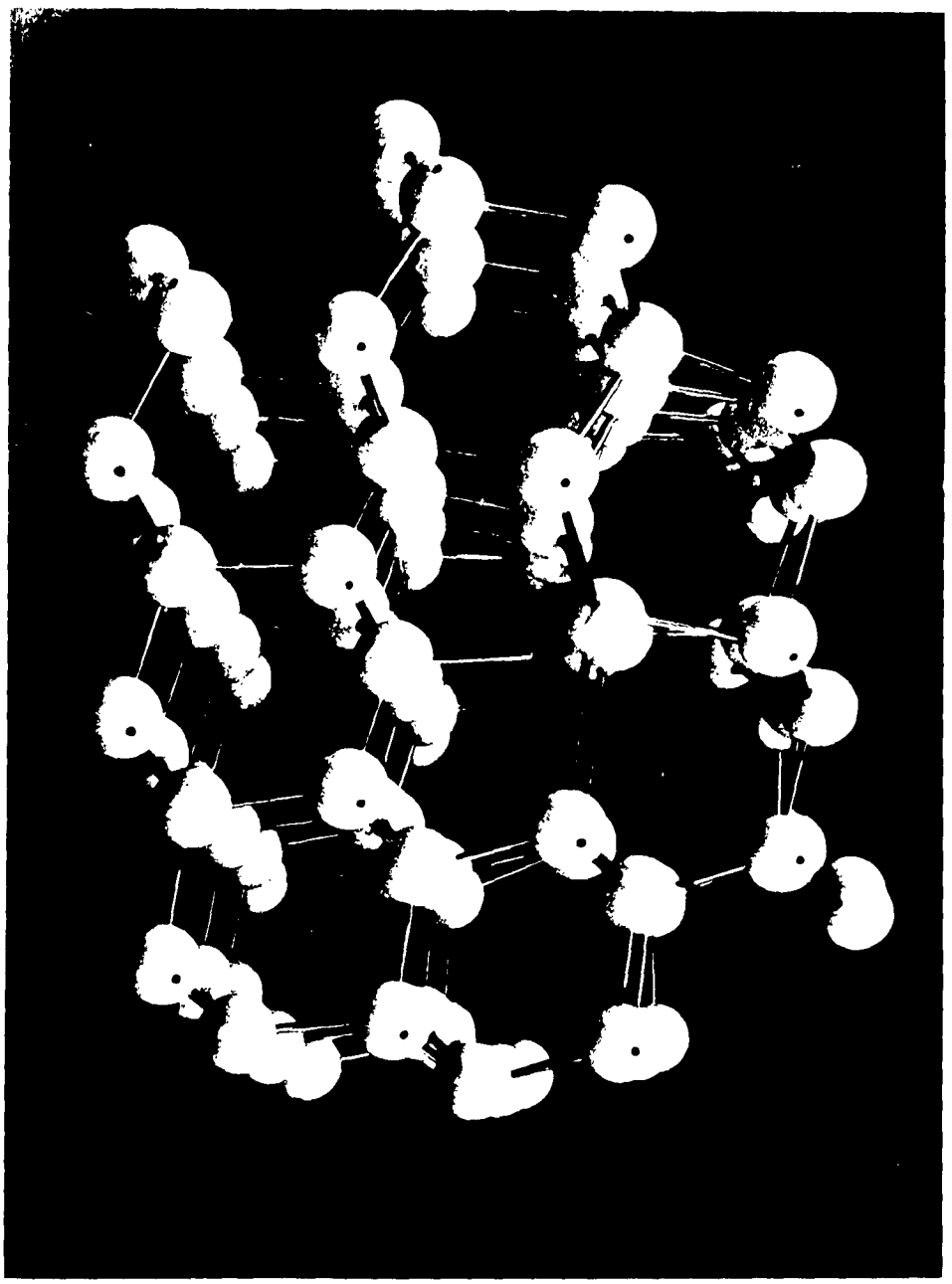
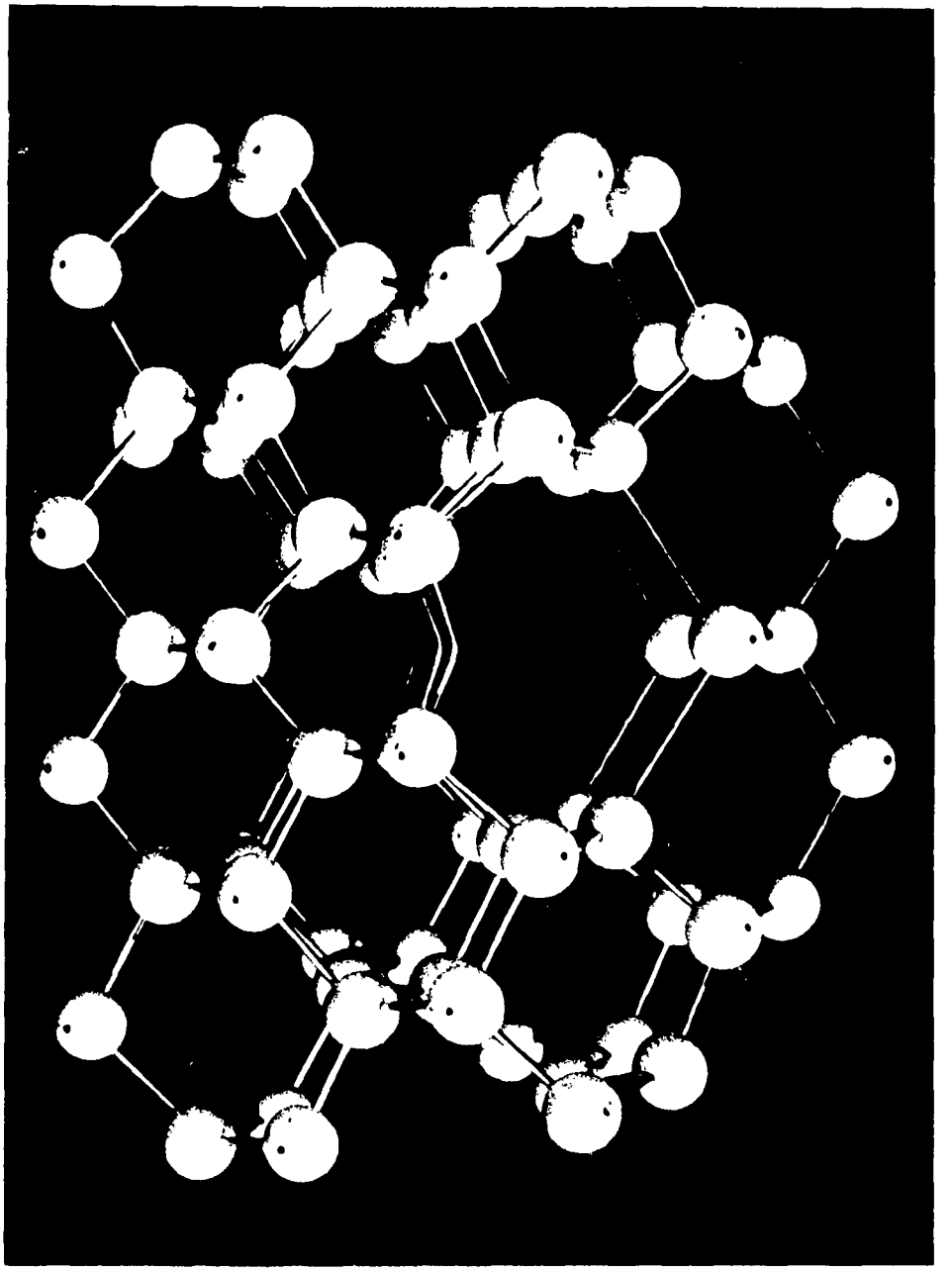


Figure 1



X-100-168

Figure 2



X-100-166

Figure 3

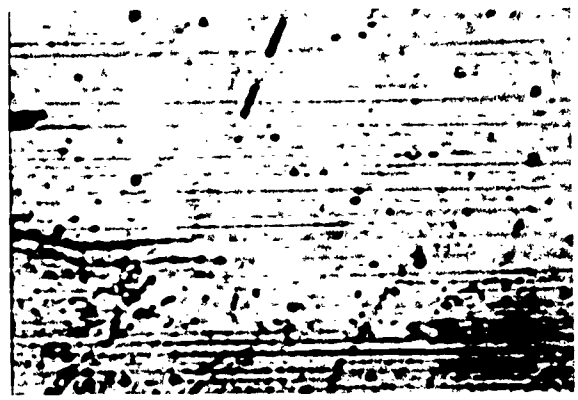


Figure 4

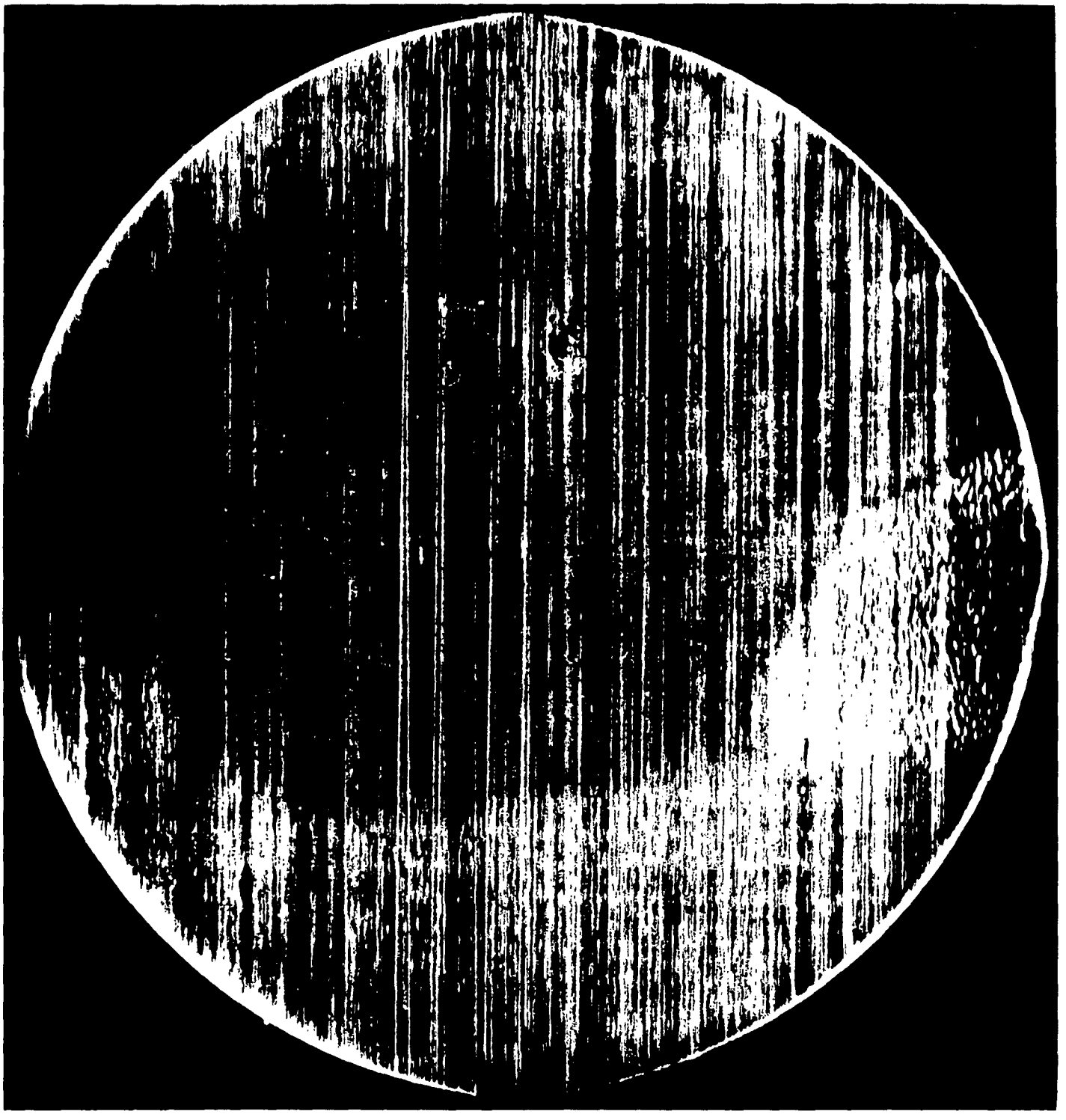


Figure 5

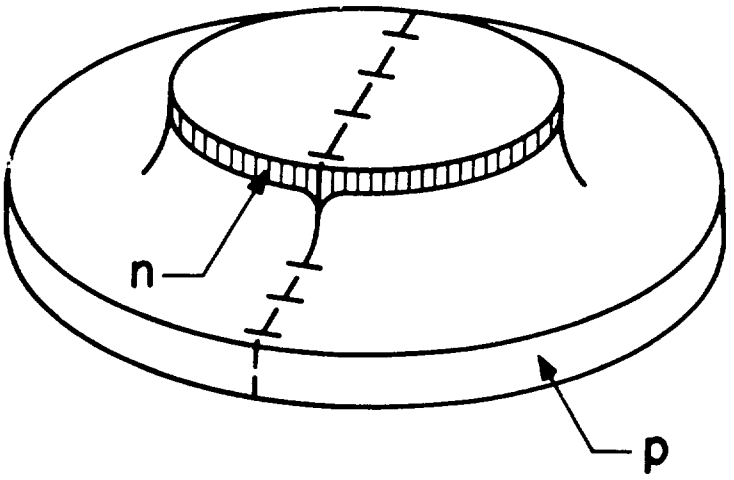


Figure 6

6-4-62
50258

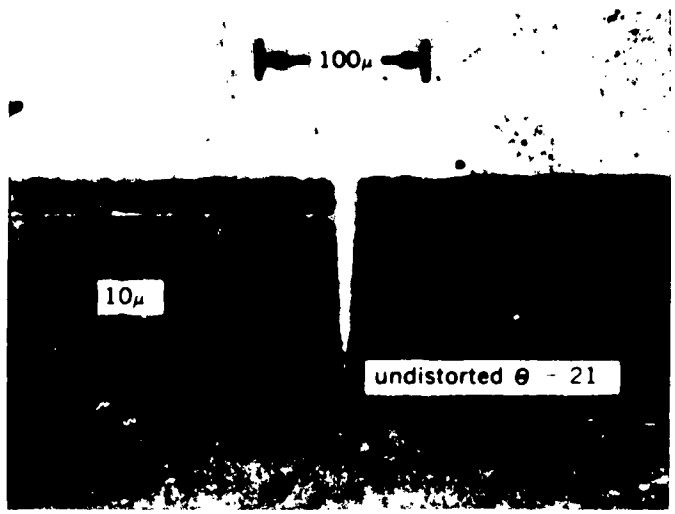


Figure 7

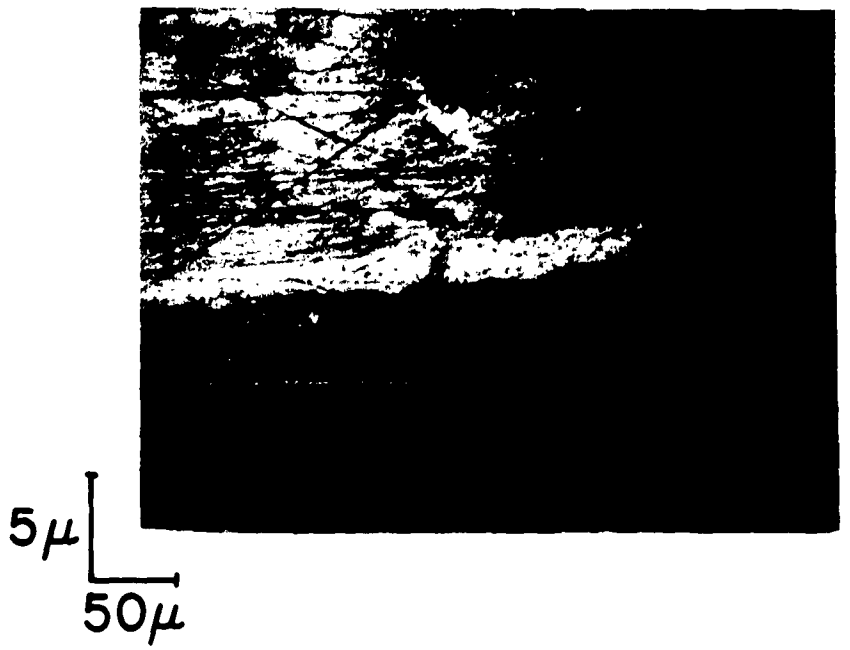


Figure 8

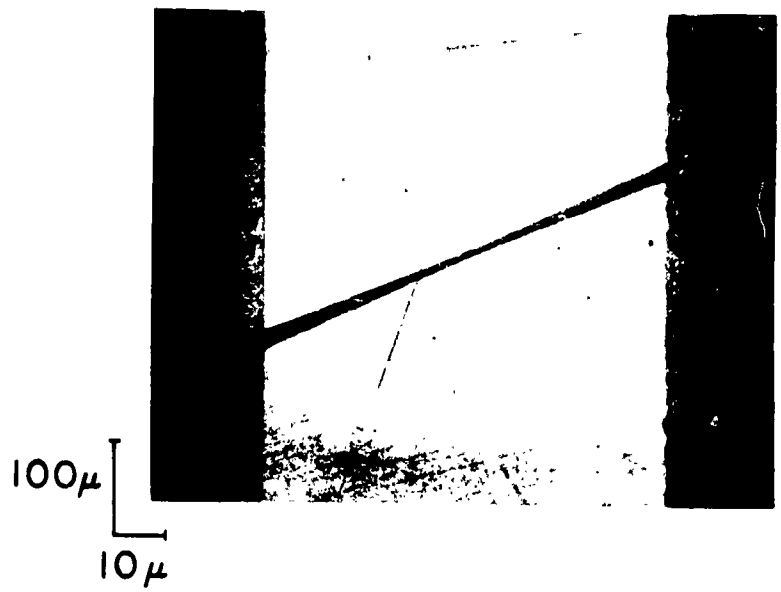


Figure 9

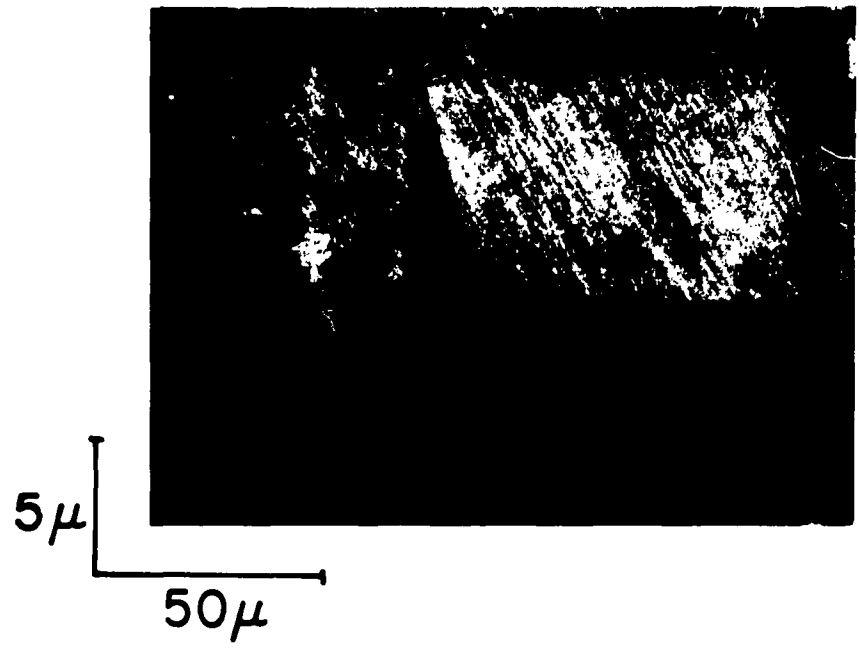


Figure 10

DISTRIBUTION LIST

	Nr of Copies
**RADC (RASGP, ATTN: Mr. A. L. Tamburrino) Griffiss AFB NY	25
* RADC (RAAPT) Griffiss AFB NY	1
* RADC (RAALD) Griffiss AFB NY	1
* GEEIA (ROZMCAT) Griffiss AFB NY	1
* RADC (RAIS, ATTN: Mr. Malloy) Griffiss AFB NY	1
* Signal Corps Liaison Officer RADC (RAOL, ATTN: Maj Norton) Griffiss AFB NY	1
* AUL (3T) Maxwell AFB Ala	1
ASD (ASAPRD) Wright-Patterson AFB Ohio	1
Chief, Naval Research Lab ATTN: Code 2027 Wash 25 DC	1
National Aeronautics & Space Admin Langley Research Center Langely Station Hampton Virginia ATTN: Librarian	1
AFSC (SCSE) Andrews AFB Wash 25 DC	1

* Mandatory

**Project Engineer will enter his symbol and name in the space provided.

*ASTIA (TISIA-2) Arlington Hall Station Arlington 12 Va	(If not releasable to ASTIA, IAW AFR 205-43, send the 10 copies to RADC (RAAPP-2) for secondary distribution)	Minimum of 10 copies
AFSC (SCFRE) Andrews AFB Wash 25 DC		1
Hq USAF (AFCOA) Wash 25 DC		1
AFOSR (SRAS/Dr. G. R. Eber) Holloman AFB N. Mex		1
Commander Naval Missile Center Tech Library (Code NO 3022) Pt Mugu Calif		1
Commandant Armed Forces Staff College (library) Norfolk 11 Va		1
Commanding General White Sands Missile Range New Mexico ATTN: Technical Library		1
ESD (ESRL) L G Hanscom Fld Bedford Mass		1
Commanding Officer & Director US Navy Electronics Lab (LIB) San Diego 52 Calif		1

*Mandatory

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mecnanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

AD ______________

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

FAILURE MECHANISMS IN SILICON SEMICONDUCTORS,
H. J. Queisser

Final Report, 1 October 1962 Contract No. AF 30(602)-2556
Covering 7-1-61 to 10-1-62 Pages, incl. illus.

This Final Report is divided into two major parts. The first part covers investigations of lattice defects in silicon and their effects on performance and reliability of semiconductor devices. Grain boundary dislocations, stacking faults in epitaxial layers, and twin boundaries are investigated. Gallium diffusion along grain boundaries is studied. Impurity atmospheres are found to influence the electrical properties of dislocations. A new method of bicrystal growing is described. Stacking faults originate from oxygen impurities at the substrate used for epitaxy. The faults cause preferential microplasma breakdown. The second part of the report treats the problem of thermal lateral instability in transistor and thermistor structures. Experimental evidence is presented for the existence of localized "hot spots" with elevated current densities and temperatures. This phenomenon is related to the "second breakdown"-failure mechanism in transistors.

Unclassified

1. Failure Mechanics
2. Semiconductors
3. Electrical Properties
4. Crystal Structure

I. Title Entry - Silicon Semiconductors
II. H. J. Queisser
III. Rome Air Development Center, AFSC
IV. Contract AF 30(602) 2556

Uniterms

Silicon
Grains
Transistors

ERRATA SHEET

The attached notice is to be inserted following the front cover of the Final Report, "Failure Mechanisms in Silicon Semiconductors," RADC-TDR-62-533, 15 January 1963, Contract AF 30(602)-2556.

SHOCKLEY TRANSISTOR
Unit of Clevite Transistor
Stanford Industrial Park
Palo Alto, California